Vol. XXX
No.

PSYCHOLOGICAL REVIEW PUBLICATIONS

Whole No. 134
1921

Psychological Monographs

A Tentative Standardization of a Hard Opposites Test

BY

MARIE HACKL MEANS, PH.D.

PSYCHOLOGICAL REVIEW COMPANY
PRINCETON, N. J.
AND LANCASTER, PA.

AGENTS: G. E. STECHERT & CO., LONDON (2 Star Yard, Carey St., W. C.)
PARIS (16, rue de Condé)

Grateful acknowledgment is due Dr. Edward K. Strong who suggested the problem and Dr. Joseph Peterson under whose guidance and help the work was completed. The writer also wishes to thank those students of Peabody College who acted as subjects, as well as the following men and women who so kindly gave of their time to make the completion of this research possible: Dr. Harvey Carr, Miss Cora Bratton, Dr. George Arps, Miss Mary Small, Dr. Walter Hunter, Dr. Florence Richardson, Dr. Benjamin Simpson, Dr. Norman Triplett, Dr. Franklin Smith, Dr. Homer Bean, Dr. E. A. Gamble, Mr. E. A. Kirkpatrick, Dr. C. A. Ruckmick, Dr. M. R. Trabue, Dr. George W. Camp, Dr. W. B. Pillsbury.

It would be most ungrateful for the writer not to acknowledge the help and encouragement received from her mother.

CONTENTS

A TENTATIVE STANDARDIZATION OF A HARD OPPOSITES TEST

I.

The Problem and its Importance

There have been many attempts to apply mental tests to college students, but the results have not been on the whole satisfactory. The very homogeneity of the college group, brought about by the natural process of the elimination of the unfit, should have warned us not to expect results comparable to those obtained from children and unselected adults. The minute individual differences of degrees of accuracy or rate of performance can be detected only by the use of a most delicate scale.

Therefore to raise our correlations of mental tests and estimated intelligence it is necessary that greater attention should be paid to the selection of the tests themselves. Where the higher functions are involved it is better to avoid tests of co-ordination and sense discrimination as well as those tests, success in which is largely dependent upon speed of performance rather than accuracy. Unfortunately the humbler task of standardization for convenient use has been neglected and until investigators have at their disposal a number of such tests, our attitude toward the validity of mental testing of college students should be highly tentative.

This present investigation was attempted in the hope of contributing a small amount toward the preparation of some adequately standardized tests. To evolve out of a test in common use a vastly more difficult test, to determine the relative difficulty of its parts, to standardize the responses which are acceptable, and to set up workable norms with which comparisons could be made—this specifically was our aim. For this purpose we have chosen an opposites test. No effort has been made to obtain correlations with other tests and no claim is made that the test, as it now stands, has mentally diagnostic values of high importance for the individual.

F.P.

II.

The Results Obtained by Previous Investigators

The diversified use of the opposites test, with its different methods of presenting and scoring, has yielded results which, while not altogether comparable, are not without interest.

Dr. Bonser[1] in an investigation of the reasoning ability of children in the upper division of the fourth, fifth, and sixth grades of the public schools of Passaic, New Jersey, tested three hundred and eighty-five boys and three hundred and seventy-two girls within the period of February 13 to 27, 1906. The tests employed were designed to exercise the most fundamental phases of reasoning ability, namely, mathematical judgment, controlled association, selective judgment, and the analytic and synthetic thinking necessary for the intellectual interpretation of literature.

For controlled association, three types of tests were used. First, in two sets of ten sentences presented, a significant word was omitted from each sentence, which was to be filled in by the pupils. Second, two other sets of ten sentences were given, in each of which two significant words were placed one above the other. The pupils were instructed to draw a line through the wrong word, leaving the sentence so it would read correctly. Third, three sets of twenty words each were given to the pupils with the instructions that they were to write beside each respective word a word just its opposite in meaning.

For selective judgment two types of tests were used. The first consisted of two sets of two series each of ten reasons why some given fact is true, some of which are correct, the others incorrect. The pupils were to check the former. Second, there were given two sets of three series each, of five definitions of a given thing or term, some of which were correct, others incor-

[1] Bonser, Frederick G., "The Reasoning Ability of Children of the Fourth, Fifth, and Sixth School Grades," *Col. Univ. Cont. Ed.*, 37, 1906, 1-101.

rect or irrelevant. The pupils were to choose the right definition. In the table below these tests are indicated as V and VI respectively.

For literary interpretation the pupils were asked to write the meaning of two stanzas of poetry.

Below is shown the coefficients of correlation obtained by correlating each test with the combined score of all the tests.

TABLE 1

Bonser's Correlations of Each Test with the Totals for All Tests in Reasoning Ability.

Opposites and Combined Score	.85
Selective Judgment and Combined Score (V)	.73
Mathematical Judgment and Combined Score	.59
Selective Judgment and Combined Score (VI)	.58
Controlled Association and Combined Score	.55
Interpreting Poems and Combined Score	.37
Spelling and Combined Score	.22

Dr. Bonser has arranged for comparison separate tables for the younger and the older group in each grade. He finds the younger group superior in the opposites test for every grade, but this superiority diminishes as the years in school increase. This would indicate, he thinks, that the test is one which reveals some sort of native ability, which is concealed in the upper grades where the test is relatively simple.

Miss Norsworthy[2] in a study of the comparison of defective and normal children found that the defectives were farthest removed from normal children in ability to deal with abstract data. Of the 137 cases ranging in age from eight years up, none of the defectives surpassed the median score for normal children in the opposites tests. This is shown in the following table.

Woolley and Fischer[3] in their work in connection with the industrially employed children in Cincinnati obtained results by the use of opposites tests which are highly significant. Over eight hundred children fourteen years of age were given the following

[2] Norsworthy, Naomi, "The Psychology of Mentally Deficient Children," New York, Columbia University, 1906.

[3] Woolley, Helen Thompson, and Fischer, Charlotte Rust, "Mental and Physical Measurements of Working Children," *Psychol. Monog.*, 1914, XVIII, 77, 213-241.

TABLE II

Norsworthy's Percentages of Children Scoring above —2 P. E., —1 P. E., and Median for Normal Children in Series of Tests.

	% above Med. for ordinary children	% above—1 P. E. (or lowest 25% of ordinary children)	% above —2 P. E. (or lowest 9% of ordinary children)
Height	45	61	77
Weight	44	66	77
Pulse	49	69	86
Temperature	26	59	77
Weight Test	18	28	39
A-T Test	1	14	28
Memory of Unrelated Words	6	18	27
Dictation	10	10	21
Memory of Unrelated Words	5	19	30
Part-Whole Test	9	17	27
Genus-Species Test	9	16	17
First Opposite Test	0	0.9	5
Second Opposite Test	0	1	7

physical and mental tests upon their entrance into the industrial world: Height, Weight, Visual Acuity, Auditory Acuity, Vital Capacity, Strength of Hand, Steadiness of Hand, Tapping, Card Sorting, Cancellation Test, Memory (Digits), Substitution, Completion of Sentences, Association by Opposites, and Puzzle Box Test.

A year later six hundred and seventy-nine of these boys and girls were re-tested.

In giving the opposites test one of the eight lists of twenty words printed one under another was presented to the subject who was requested to write beside each word another word opposite in meaning. The time was recorded for the total list but the scores were based on the percentage of accuracy alone. Credits of one, one-half, or zero were assigned to the responses given. Misspelled words were given full credit but adjectives written in place of adverbs received only half credit.

A positive correlation was obtained between the school grade completed at fourteen years of age and the ranking in every one of the mental tests. The general order is as follows: Memory,

Association by Opposites, Sentence Test, Substitution Test, Cancellation, and Puzzle Box.

The public school children were found to be superior to the parochial school group, and their superiority was most decided in opposites and in the puzzle box tests, the two tests which, according to the investigators, are farthest removed from a relationship with school drill.

"There is a marked and consistent positive correlation with school grade in this test (opposites) for both sexes and at both ages. The differences from grade to grade are so large and so consistent that their significance cannot be questioned."[4]

While director of the Department of Psychology at the New York State Reformatory for Women, at Bedford Hills, Dr. Weidensall[5] began a series of experiments upon the women detained in that institution. It was hoped that a number of tests might be found which would prove prophetic of the convicted woman's reformability and would thus eliminate such cases which, because of inability to learn, consumed a disproportionate amount of time.

The major portion of the monograph deals with the records of the criminal woman as compared with the norms of normal working girls fourteen and fifteen years of age. The norms were being formulated at the time by Dr. Woolley[6] in her work as director of the Bureau of Vocational Guidance, Cincinnati, Ohio. Since comparisons were to be made, the tests which were being used by Dr. Woolley were adopted by Dr. Weidensall and given in approximately the same manner.

The women tested were admitted between the first of January and the end of October, 1913. Two hundred and eight were committed to Bedford during that period, but of that number only a hundred were tested. Because of lack of facility in the English language of twelve foreigners, the percentile tables and curves are based on eighty-eight records. All tests were given

[4] Woolley, Helen Thompson, and Fischer, Charlotte Rust, *Op. cit.*, 222.

[5] Weidensall, Jean, "The Mentality of the Criminal Woman," Baltimore, Warwick & York, 1916, 3-266.

[6] Woolley, Helen Thompson, and Fischer, Charlotte Rust, *Op. cit.*

individually and during the first two weeks of confinement while the subject was still in quarantine.

The method followed in connection with the opposites test was to present a list of twenty words and to record the total time required for the subject to write the opposites. Results were based on the percent of accuracy, as it was found that the correlation between rank in time and accuracy in this test was as high as .83, P. E. .029.

All identical accuracy scores were arranged in point of time of performance and the rankings correlated with the native ability of the group as estimated by the director of the Industrial School of the Reformatory after she had worked with these women for from eighteen months to two years. This correlation was + .79, P. E. .026, and was obtained by the formula

$$r = 1 - \frac{6 \Sigma d^2}{n (n^2 - 1)}$$

This correlation, which was higher than that for any other test, would probably have been even higher had it not been necessary for the director to base her judgment of ten of the women on the reports of matrons. It was the difference in rank accorded these ten which was responsible for some of the largest variations from the rank of the test.

It is interesting to note that of the Bedford eighty-eight but thirty-nine percent attain or surpass the median record of the working girl of fifteen. Dr. Weidensall feels that the easy opposites test proved the most reliable of all tests given, for clinical purposes.

In 1912 Dr. Benjamin R. Simpson[7] selected two groups as widely different in intellectual status as possible, the one represented by seventeen professors and advanced students in Columbia University, the other by inmates of charitable institutions, with the exception of two who were recognized by their associates as being dull. The fifteen tests, which were given, were administered individually and in the same order.

The hard opposites test not only separated the two groups com-

[7] Simpson, Benjamin R., "Correlations of Mental Abilities," *Col. Univ. Cont. Ed.*, 1912, No. 53.

pletely but correlated with the general intelligence of the good group, as estimated by the members which composed it, more highly than any of the other tests. The individuals of the good group were rated in order of merit for general intelligence, each by the rest of the group, four years after the tests were given. Two rankings made by the experimenter a month apart were included. The judgments, correlated with the various tests, are as follows:

TABLE III

Simpson's Correlations of the Estimated Intelligence of his Superior Group with the Results of Eleven Tests

Estimated	Intelligence	and	Hard Opposites	.96
"	"	"	Ebbinghaus Completion Test	.89
"	"	"	Memory for Words	.93
"	"	"	Memory of Passages	.35
"	"	"	Easy Opposites	.82
"	"	"	Adding	.72
"	"	"	Learning Pairs	.34
"	"	"	Completing Words	1.00*
"	"	"	"A" Test	.21
"	"	"	Geometrical Forms	.07
"	"	"	Drawing Lengths	—.20

*This coefficient, according to Dr. Simpson, is not to be considered reliable, since the reliability coefficients of the Completing Words test in the Good group is only .27.

On account of the high correlation between the hard opposites and the Ebbinghaus Completion Test, Dr. Simpson feels that they test the same mental function, namely selective thinking.

Bronner[8] has endeavored to determine the intellectual status of the delinquent girl as compared with the intelligence of several other groups engaged in occupations and pursuits which, of necessity, required varied degrees of education and ability. Thirty girls living at Waverly House, a detention home maintained by the New York Probation Association, composed the delinquent group. The college group contained thirty-six girls, all members of the freshmen and sophomore classes of Barnard and

[8] Bronner, Augusta F., "A Comparative Study of the Intelligence of Delinquent Girls," *Col. Univ. Cont. Educ.*, 1914, No. 68.

Teachers' College. Thirty-four Jewish girls, who spent their evenings at the University Settlement or the Harlem Branch of the Y. W. C. A., engaging in educational pursuits, formed the evening class group. The fourth group was composed of twenty-nine girls who had never engaged in any wage-earning occupation except domestic service. None of these girls was pursuing studies whereby she hoped to prepare herself for a different occupation nor had she been guilty of offenses which had brought her in conflict with the law.

In addition to an ethical discrimination test, those tests employed comprise the Easy Opposites, the Hard Opposites, the Memory of Words, the Memory of Passages, and the Ebbinghaus Completion test. These tests were the same as those used by Dr. Benjamin Simpson in his study of *Correlations of Mental Abilities.* In the following table, which is a reorganization of five tables presented by Bronner, D represents the delinquent group, C the college group, E the evening class group, and S the domestic service group. These data show, as Bronner states, that the

TABLE IV

Comparisons of the Four Groups

Easy Opposites	
% of D reaching the 25 percentile of the C group	3.3%
% of E reaching the 25 percentile of the C group	15.0%
% of S reaching the 25 percentile of the C group	7.0%
Hard Opposites	
% of D reaching the 25 percentile of the C group	0.0%
% of E reaching the 25 percentile of the C group	12.0%
% of S reaching the 25 percentile of the C group	0.0%
Memory of Unrelated Words	
% of D reaching the 25 percentile of the C group	20.0%
% of E reaching the 25 percentile of the C group	59.0%
% of S reaching the 25 percentile of the C group	7.0%
Memory of Passages	
% of D reaching the 25 percentile of the C group	10.0%
% of E reaching the 25 percentile of the C group	15.0%
% of S reaching the 25 percentile of the C group	3.0%
Ebbinghaus Completion Test	
% of D reaching the 25 percentile of the C group	6.6%
% of E reaching the 25 percentile of the C group	9.0%
% of S reaching the 25 percentile of the C group	3.0%

college girls excel in all the tests. The delinquent girl is quite as capable as the domestic service girl, while both are surpassed by the girls attending evening classes.

The superiority of the evening class girls over the other two groups displayed in each test, cannot be attributed to more favorable educational advantages for none of the members had attended high school, whereas of the delinquent group, four had reached the eighth grade, two the first year of high school, two the second year, one was graduated, and one had attended a normal school.

The Hard Opposites Test separated the college group from the others almost entirely. Next in order comes the Ebbinghaus Completion Test, with the Easy Opposites ranking third. It is not without interest that Dr. Simpson found the Hard Opposites separated his good group entirely from his poor group and the Easy Opposites in this respect surpassed the Ebbinghaus Test.

Let us now examine the record of the Opposites Test when applied to the highly selected group, represented by the college student. In 1914 Dr. Kitson[9] began a series of tests upon the freshmen in the College of Commerce and Administration of the University of Chicago. One of the sixteen tests used was an opposites test. Two lists of twenty words each were presented and the time recorded for the subject to call the opposites of each list. Five points were deducted for each wrong word or for failure to respond within fifteen seconds, from the one hundred points allowed each list. The final score for the individual was obtained by dividing the time score by the accuracy score.

The stimuli were of such little difficulty that they measured for the most part merely speed of association. Almost half of the subjects obtained an accuracy score of one hundred in each list. Of forty students, thirty-one scored one hundred in accuracy in the easy list, and twenty-one scored a hundred in the hard list.

In his results Kitson has included the following table of correlations.

Of the fifteen tests used by Kitson, by computing the multiple and the partial correlations and the regression coefficients,

[9] Kitson, H. D., "The Scientific Study of College Students," *Psychol. Monog.*, 1917, 23.

TABLE V

Correlations of Standings in Each Test with Standings in the Net Score (Method of Rank Differences)

	Correlations	P. E.
Logical Material Seen (Deferred)	.60	.07
Opposites	.53	.08
Hard Directions (Printed)	.49	.08
Objects Seen	.48	.08
Loss in Logical Material Seen	.47	.09
Logical Material Heard (Deferred)	.45	.09
Word Building	.45	.09
Loss in Logical Material Heard	.43	.09
Sentence Building	.42	.09
Constant Increment	.38	.10
Business Ingenuity	.33	.10
Logical Material Seen (Immediate)	.29	.10
Numbers Heard	.27	.10
Hard Directions (Oral)	.23	.11
Logical Material Heard (Immediate)	.23	.11
Number Checking	.18	.11

Rosenow[10] concludes that five of the tests carry all the meaning with reference to school marks and hence all diagnostic value. In the table below the writer has rearranged the results of Rosenow's investigation. From it one learns that the probability is 1300 : 1 that the Logical Memory is significant.

TABLE VI

Rearrangement of the Results of Rosenow's Investigation

Probability is 1300 : 1 that the Logical Memory Test is Significant.
" " 25 : 1 that the Constant Increment Test is Significant.
" " 30 : 1 that the Sentence Building Test is Significant.
" " 23 : 1 that Auditory Presentation is Superior to Visual.
" " 140 : 1 that the Loss or Gain in Logical Auditory Memory is significant.
" " that the Hard Directions Test has Negative Significance.
" " that the Objects Seen Test has Negative Significance.

The remaining tests, including the Opposites, have no diagnostic value.

In view of the fact that the stimuli used in the Opposites Test were the Woodworth and Wells standardized list, Rosenow's con-

[10] Rosenow, Curt, "The Analysis of Mental Functions," *Psychol. Monog.*, 1917, 24.

clusions are not surprising. The words were of so little difficulty that they tested only speed of association.

King and M'Crory[11] in testing two hundred and seventy-six women and two hundred and sixty-eight men in the freshmen class at the University of Iowa, applied seven tests, including both easy and hard opposites. The following table indicates the correlations which were obtained between the Opposites and other tests as well as university grades. The Opposites Test, they find, correlates most highly of any of the tests with the university marks of both men and women.

TABLE VII

Correlations of Opposites with Other Tests and University Grades

	Correlations	
	Women	Men
Opposites and Completion	.31	.79
" " Arithmetic (Speed)	.03	Neg.
" " Arithmetic (Accuracy)	.01	Neg.
" " Analogies	.52	.77
" " Information	.24	.56
" " Visual Imagery	.07	.56
" " Logical Memory	.32	.38
" " Test Average	.51	.88
" " University Grades	.45	.84

Dr. King[12] is of the opinion that the Opposites Test, if thoroughly standardized and used in conjunction with other tests, will yield results of great importance.

In the Spring of 1916, while the writer was yet engaged in the selection of appropriate stimuli for the present investigation, an opportunity[13] presented itself of correlating the grades of seventy-three students in psychology with the results secured by the use of the Opposites Test. An effort had been made to keep the two lists composed of one hundred and fifty words each, as nearly

[11] King, Irving, and M'Crory, J. L., "Freshmen Tests at the State University of Iowa," *Jour. Educ. Psychol.*, 1918, IX, 32-46.

[12] King, Irving, and Gold, Hugo, "A Tentative Standardization of Certain Opposites Tests," *Jour. Educ. Psychol.*, 1916, VII, 459-482.

[13] Dr. Edward K. Strong, at that time Professor of Psychology at George Peabody College for Teachers, kindly furnished these data. The grades were compiled from the records of fourteen tests given during the quarter.

equal as possible. The two lists are given below. The words are arranged in random order. Those words which had not been used by previous investigators are in italics.

List I.—Defective, Late, Sinful, Easy, Hilly, Superior, Girl, White, Cool, Large, Evil, Queen, Deep, True, Public, Sink, Future, Adroit, Dangerous, Day, Ugly, Quick, Poor, Diligent, Wicked, Round, Ceiling, Broken, Gentle, Vague, Brief, Animated, Slovenly, Dim, Out, Rude, Lazy, Injurious, Conservative, Wet, Asleep, Stingy, Fertile, Wise, Calm, Tardy, Hinder, Respect, Big, Gain, Great, Profit, Young, Few, Summer, Above, Glad, Masculine, Remember, Off, Beginning, Love, Straight, War, Joy, Naked, Pride, Apart, Brave, Noisy, Fickle, Create, Wild, Despondent, Frequently, Timid, Hollow, Belief, Bad, Up, Sick, Empty, Strong, Inside, Front, After, Broad, Sharp, Sweet, Succeed, Add, Happy, Raise, Aristocratic, East, Short, Thick, Result, Rare, Stale, North, Hostile, Laugh, Obnoxious, Expensive, Near, Join, Hot, Forcible, Preserve, Strict, Handsome, Friend, Miser, Exciting, Rough, Brother, Light, Careful, Push, Haughty, Impoverish, Busy, Much, Graceful, Ocean, Precise, Barbarous, Ignorant, Reckless, Odd, Victorious, Repulsion, Permit, *Positive, Pessimistic, Extravagant, Durable, Analytical, Parsimony, Orthodoxy, Acute, Exoteric, Antonym, Dorsal, Longitude, Divide, Infinity, Dynamic, Posterior.*

List II.—Gay, Foolish, Drop, Giving, Cloudy, Blunt, Beautiful, Backwards, Well, Top, Success, Soft, New, Refined, Weary, Spend, Break, Male, Country, Dark, Weak, Black, Disastrous, Rigid, Elation, Hindrance, Savage, Degrade, Ripe, Shaky, Separate, Liquid, Sell, Honest, Difficult, Dirty, Wrong, Winter, Helpless, Obscure, Expand, Insignificant, Sleepy, Sad, Little, Enemy, Open, Yours, Yes, Conservative, Soothing, Doubtful, Sacred, Sure, Reveal, Stupid, Motion, Sickly, Slowness, Outside, Same, Cowardly, Float, Foreign, Strength, Sane, Level, Simple, Many, Lost, Something, Sour, Enrage, Serious, Long, In, Take, Tight, Prompt, Patient, Permanent, Genuine, Morning, Smooth, Heavy, Full, Grand, Humility, Tall, Over, First, Strife, Follower, Hold, Proficient, Vertical, Shallow, Absent, Rapid, Rich, Purity, Loquacious, Imaginary, Silly, Increase, Wider, Nowhere, Upper,

Woman, Generous, Careless, Scarce, Height, Always, Wife, Best, Below, Thin, High, Early, Suave, Lack, Advance, Harmonious, Bless, Both, Cruel, Cheap, Ancient, Less, Forget, Come, Slow, Good, *Negative, Optimistic, Economical, Perishable, Synthetical, Prodigality, Heterodoxy, Chronic, Esoteric, Synonym, Ventral, Latitude, Multiply, Zero, Static, Anterior.*

The test was given individually and orally. The experimenter read the stimulus and recorded the response of the subjects as well as his reaction time. The accuracy score in percentage and the time score arrived at by computing the median time of all reactions were tabulated for each student.

In grading these papers, credits of one, one-half, or zero were assigned to the responses. What credit a word deserved was determined solely by the writer, as she was not yet ready to standardize the responses for so bulky a list.

In order to allow equal credit for time and accuracy the following formula was used:

$$\text{Individual score} = \frac{\frac{Dt}{QT} + \frac{Da}{QA}}{2} \text{ in which}$$

Da = the deviation of the individual accuracy scores from the accuracy scores of the group.

Dt = the deviation of the individual time score from the time score of the group. Since signs were regarded, where an individual had a lower time score than the group his deviation was positive.

QA = the quartile of the accuracy scores of the group.

QT = the quartile of the time scores of the group.

The coefficient of correlation obtained by the method of rank differences was .54, P. E. = .08. Speed and accuracy were likewise correlated, giving a coefficient of .33, P. E. = .10. This coefficient would hardly justify the statement that, in so far as this particular test is concerned, either accuracy or speed could be disregaded without affecting the rankings in the test.

In Figure I is shown the relative positions of the several members of the class in psychology and in the Opposites Test. The

vertical line indicates the rank in the test, the horizontal line, the rank in psychology. A glance at the figure will show that on the whole those good in the test were good in psychology. The greatest exception is to be found in the position of the student indicated by the figure 5 in the graph. This woman of splendid ability was permitted by her physician to remain in school only on the condition that she would do the least possible amount of work to secure credit for the course in which she was enrolled. Case 4, a widow, had been out of school and school work for a number of years. The increased cost of living had forced her to return. Case 3 had registered in psychology because it was one of the required courses.

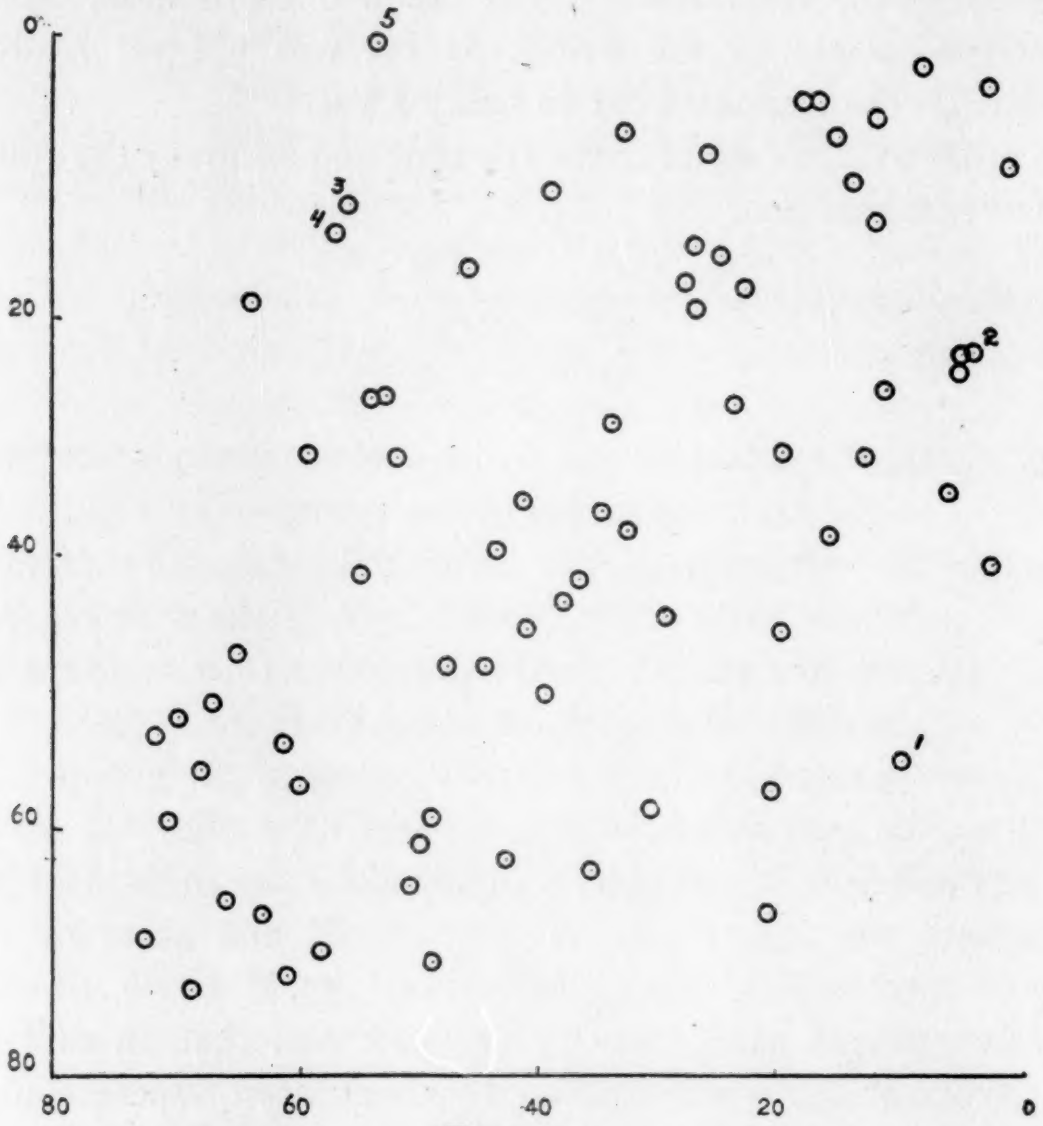

FIGURE 1

The relations between rank based on psychology grades and rank based on the Opposites Test score are shown here. Each circle indicates the position of one of the seventy-three college students. Units on the vertical line denote rank in the test; horizontal units are ranks in psychology.

On the other hand, it seems, one could predict with less certainty the standing in the class of those students who are somewhat poor in the test. This may be due in part to the inadequacy of the test itself, in part to the inequality of interest in psychology. Cases 1 and 2 are typical of those students whose class work represents their best effort. These cases are selected because they are the only ones falling within the writer's experience. A careful analysis of the individual records, if possible, might reveal results of significance.

While the results indicate that the test as used by the different investigators reveals some sort of native ability, just what this ability is and to what extent it is revealed remains to be seen. Certainly, results of various experimenters up to date are not comparable, nor will they be until identically the same test is given with the same method of presenting and scoring.

III.

Previous Efforts toward Standardization

The first attempt to meet this need of standardization was made by Woodworth and Wells[1] in 1911. After testing six subjects with a long list of words, forty words were selected, which in turn were tried with forty other subjects. A few more words were substituted from tests of a few subjects, then the revised collection was tried with thirteen fresh subjects (all college and graduate students) and a few more minor corrections introduced.

In the experiment the four lists of ten stimulus words were presented visually but the time of the single reaction as well as the total time was roughly taken.

An attempt was made to have the halves of the list of equal difficulty, and since it was found impossible to prepare a list of twenty stimuli of equal difficulty, the words were combined in pairs, "so that pairs should be of equal difficulty, as judged by the sum of the reaction times to the two members of each pair." One pair, for instance, consisted of the hardest and the easiest word in the list and another pair of two words of medium difficulty.

In the event that the test might be given with a time limit, the authors have placed the words of medium difficulty in the list where most of the subjects will be stopped, namely, from about the eighth to the sixteenth word. "If then the time limit is so chosen that the great majority of subjects shall be stopped in this list, the separate words may, without much error on the average, be counted as equal units."

The test as given by Woodworth and Wells is scored only on the basis of time. The words were so selected that none of the subjects could fail to have a perfect accuracy score. As such, it is, as the authors intended, merely a test of the speed of association.

[1]Woodworth, R. S., and Wells, F. L., "Association Tests," *Psychol. Monog.*, 1911, XIII.

In 1916 a further attempt was made by King and Gold[2] to standardize the test. The four lists of twenty easy words and the four lists of twenty hard words, as used by Simpson,[3] were presented individually to nine faculty members, twenty-three graduare students, forty-seven seniors, and twenty-one juniors of the Departments of Education and Psychology in the University of Iowa.

The subjects were instructed to take one list at a time, go down the list and name orally the opposite of each word in rapid succession and to avoid wherever possible prefixing a syllable to the stimulus word. The time was recorded for each list separately and the responses were taken down in shorthand.

The easy and hard opposites were scored differently. In the case of the easy opposites a credit of one, one-half, or zero was given to responses, and the penalty for an omitted word was four seconds. To the various responses to the hard opposites, credits of one, two-thirds, one-third, and zero were assigned, and an omitted word or incorrect word was penalized eighteen seconds. Regardless of the fact that the words were of varying difficulty, the penalty remained the same, namely, four seconds for easy words and eighteen seconds for hard words.

On the basis of these results (one hundred records for each word) the percent of failures for each word was computed. For each stimulus is recorded the value in terms of accuracy percentage as well as the acceptable responses and the frequency of each.

In the work of Harry A. Greene[4] we find the first attempt to assign to each word a point value based on its relative difficulty. Greene presented the stimuli used by King and Gold[5] to two groups of freshmen in the University of Iowa in the fall of 1916-1917. Nine hundred and ninety individuals were tested by the first half of the list and seven hundred and ten by the second

[2] King, Irving, and Gold, Hugo, "A Tentative Standardization of Certain Opposites Tests," *Jour. Educ. Psychol.,* 1916, VII, 459-482.

[3] Simpson, Benjamin R., *Op. cit.*

[4] Greene, Harry A., "A Standardization of Certain Opposites Tests," *Jour. Educ. Psychol.,* 1918, IX, 559-566.

[5] King, Irving, and Gold, Hugo, *Op. cit.*

half. The words were graded by the King and Gold standardization with the exception that words given two-thirds credit by them were given one-half credit and those receiving one-third credit were scored wrong.

The percentage of failures was then determined by allowing a value of one unit for correct responses and one-half for half-correct responses and subtracting this total from nine hundred and ninety. In the case of those words from the last half of the list for which there were only seven hundred and ten responses available, these totals were increased in the proportion of nine hundred and ninety to seven hundred and ten, thus making all the words in the list comparable.

By reading directly from the table based upon the area of the probability curve and assuming that the base line is broken arbitrarily at +3 sigma, the percentage scores were changed into percentile values. These values were then totaled and each value in turn divided by the total, thus converting the percentile values into relative point values. The points were based on accuracy alone, no account being taken of time.

IV.

The Selection of Stimuli for the Present Investigation

At the very outset, the writer was confronted with the problem of the method in which the test was to be presented. Chiefly because of the opportunity it afforded of studying each word individually, the oral method was chosen. Each word thus became a unit. The list could be lengthened or shortened at will without influencing to any appreciable extent the words retained. Let it be held in mind that at this point our interest was two-fold, namely, the selection of suitable stimuli and the computation of the value in points to be assigned to each word.

To have given the test in groups would have simplified the scoring and would have dispensed with much effort, but the attendent disadvantages made it necessary to abandon this method as a possible choice.

In the first place, it was intended that the test should be more difficult than previous tests of this character. This meant the substitution of entirely new stimuli, stimuli to be acquired only by the slow process of trial and error. Furthermore it was encumbent upon the writer, due to the small number of subjects at her disposal, to avoid a method which would involve discarding the entire record of a subject because of the use of one or more undesirable stimuli. As given, each word was a problem apart from every other word and only the records of those words not in the list as finally decided upon, were thrown into the discard.

If the group method is to be employed, a decision must be made as to whether time, accuracy, or a combination of both will be considered a measure of the difficulty of the word. Time considered alone, the test becomes one largely of speed of association. If the test is to be given with a time limit, the individual responses lose their identity as units and the percentage of failures

per stimulus will vary with the time allowed. For instance, if the time limit is reduced from six to three minutes, the percentage of failures will vary considerably, particularly in the case of the more difficult words which should be placed at the end of the list. Thus one can see how a shortened time limit might give identical accuracy scores with words of unequal difficulty.

To illustrate the effect of varying the time limit not only upon the percentage of failures but also upon the nature of the responses themselves, let us consider the responses to the word "sacred" given, first, by a group of fifty-six students tested orally and individually, allowing ten seconds for the response, second, by another group of fifty-six students similarly tested, and third, by a group of fifty-six who were allowed only six minutes to write the opposites to the list of sixty-eight words eventually retained. In the table below is shown the number of failures and the exact responses given by the three groups of fifty-six. Notice the greater number of failures in Group III and also the homogeneity of responses. This may be explained by the fact that the word is not reached in the list or that uncertain words are passed up for those in which the subject feels confident of success.

On the other hand to lengthen the time limit considerably, while it would give more nearly correct accuracy scores, would be unfair to the individuals tested, if these same records were to be used as the basis for norms. With a lengthened time limit, students would be able to complete the test at different intervals. Evidently of two students, both of whom make perfect scores, the one finishing the work in the shorter period is the more efficient. But we should have no measure of this efficiency.

To give the test as a group test without taking into consideration the time consumed, would seem to imply either that additional time would fail to increase the accuracy percentage, or that without exception the more difficult a word, the longer the time required to think of an opposite. The former assumption is probably true within certain limits. It is conceivable that a time limit might be selected beyond which there would be no improvement, but it is obvious that the time limit might be so shortened

TABLE VIII

Showing the Different Responses Given as an Opposite to the Word "Sacred" when the Time Limit is Changed

Different Responses Given by the Groups	Number of Times Each Response is given by the Several Groups		
	Group I	Group II	Group III
Common	2	1	1
Secular	0	2	4
Profane	5	7	4
Sinful	0	2	1
Worldly	0	2	1
Unholy	9	1	2
Irreverent	2	3	1
Wicked	2	2	1
Sacrilegious	3	7	2
Unsacred	8	5	0
Vulgar	1	3	0
Public	2	1	0
Mean	0	1	0
Ungodliness	0	1	0
Heathen	0	1	0
Sacrilege	0	1	0
Obnoxious	1	0	0
Ungodly	1	0	0
Unnoticed	1	0	0
Idolatrous	1	0	0
Irreligious	1	0	0
Infidel	1	0	0
Hypocritical	1	0	0
Failure to Respond	15	16	39
Number of Different Responses per Group	16	16	9

that the accuracy scores would be greatly affected. In general it is true that words with a high accuracy score have a correspondingly low time score, but a glance at Tables XI and XII will show that such is not always the case, for words with different accuracy scores may have identical time scores and words with identical accuracy scores have different time scores. However, the correlation between the two is undoubtedly high.

To give the test individually and record the time for the total list has all the disadvantages in point of labor and none of the advantages that accrue when the test is given individually and orally.

Having determined the procedure to be followed, the writer began in the fall of 1915 the task of selecting suitable stimuli for the present investigation. After eleven records were obtained with a list containing the three hundred and twenty-three words used by previous experimenters and forty new words, the list was divided and during the remainder of the year two lists were used which consisted on the average of about one hundred and fifty words each. From time to time words found inadequate were dropped and others substituted as they occurred to the writer. Hence the inequality in the number of subjects tested with a given word. In this manner one hundred and thirty-nine students at George Peabody College for Teachers were tested during the school year 1915-1916.

Commonly misunderstood words were weeded out as were those with an accuracy score of one hundred percent. Likewise attempts were made to discard those words which had an opposite formed by adding the prefix "un" in frequent and reputable use. Meanwhile original words were subjected to the same process of examination and elimination.

The point of interest was the individual word, but as these data were to be used in another connection care was taken that the conditions remain constant throughout. Since each test was given individually, it was an easy matter to change the order of the stimuli and thus avoid practice effects. The reaction time in fifths of a second was recorded with a stop-watch along with the response. The following directions were read to each subject:

"As soon as I read a word you are to give me the best opposite you can think of. For instance, if I read 'black' you are to say 'white.' Do not give me phrases, nor words beginning with the prefix 'non.' The word you give must belong to the same part of speech as the word in the list. Your time will be recorded, so answer each word as quickly as possible. Under no circumstances will I allow you over ten seconds for a word. Let me show you how long ten seconds really is. (Experimenter illustrates with stop-watch.) Remember in each case to respond as quickly as possible. Do you understand what you are to do?"

In the fall of 1918[1] the work was resumed. It was decided expedient to use only one list as the number of possible subjects to be obtained at Peabody was small. Two lists would of course cut the records for each word in two. Besides the second list contained for the most part opposites of the first. Of the one hundred words with which the work was begun in the fall, thirty-two were dropped, leaving the sixty-eight which comprise the list as it now stands. These words, arranged in the ascending order of difficulty, are given below. The original words are given in italics. Full, *Negative,* After, Dim, Blunt, Success, *Pessimistic,* Joy, Public, Profit, Spend, Always, Graceful, Strength, Ancient, Expand, Barbarous, Hinder, Despondent, Vague, Fertile, Doubtful, Injurious, Busy, *Abstract,* Advance, Foreign, Create, Simple, *Extravagant,* Aristocratic, Rare, Dangerous, Slovenly, Defective, Stingy, Reveal, Diligent, Join, Impoverish, Permanent, Elation, Sinful, Obnoxious, Conservative, Victorious, Obscure, Proficient, Rigid, Repulsion, Imaginary, Permit, *Orthodoxy, Analytical, Extrinsic, Sacred, Dynamic, Loquacious, Heterogeneous, Spurious,* Disastrous, *Facility,* Pride, Result, Adroit, *Parsimony,* Suave, *Esoteric.*

One hundred and twelve subjects were tested with this list. This means that later in determining the relative difficulty of the stimuli, the calculations are based on at least one hundred and twelve records for each word, varying up to one hundred and eighty-seven for others.

[1] This research was conducted under the supervision of Dr. Edward K. Strong up to this point, thereafter under the guidance of Dr. Joseph Peterson.

V.

The Choice of the Acceptable Responses

The next step was to standardize and evaluate the responses. In order to facilitate matters, beneath each stimulus was written the long list of words which had been given as opposites by those subjects experimented upon. The credit due each of these responses was determined by five judges, including the writer. They were besides the writer:

Miss Lula O. Andrews, Professor of English; Miss Mary Clay Hiner, Instructor in English; Mr. S. H. Phelps, Instructor in School Administration, and Dr. Joseph Peterson, Professor of Psychology, all of George Peabody College.

Each judge was ignorant of the credit assigned any word by any other judge. The following directions, a copy of which was handed each judge, will make the matter clear:

Directions for Grading Responses

I. Make use of any available source of information such as the dictionary or book of antonyms.

II. The responses are to be graded with a grade of "one," which means an exact opposite, or "one-half," which means only an approximate opposite, or "zero," which means a failure.

III. More than one response to a word may be given a credit of "one."

IV. Words belonging to a different part of speech are to be graded "zero," also words with the prefix "non."

V. If a word belongs to the same part of speech and is nearly an opposite, give it "one-half" credit.

VI. Add any response that may suggest itself to you and grade it as directed above.

The new words suggested were in turn passed on by the other four judges. In addition, when later the test was converted into a group test and hundreds of records secured, a number of

different responses, given by those tested, were submitted to the same process of evaluation.

The credits assigned each response were then averaged and if the result was nearer "zero" than "one-half," the word received no credit; if nearer "one-half" than "zero" or "one" it received "one-half" credit; and if nearer "one" than "one-half," it was given a credit of "one." For instance the response "young" in answer to "ancient" received credits as follows:

TABLE IX

	I	II	III	IV	V	Total	Aver.	Credit Assigned
Credits given to "young" by five judges..	1	½	0	0	½	2	.4	½

On the following pages is to be found in Column II the stimulus word, in Column III the responses which are due a credit of "one' (the full value assigned to the word), and in Column IV those responses which receive only "one-half" credit (half the value assigned the word). The value of the word is given in points in Column I. The manner in which we arrived at these values will be discussed in the following section.

TABLE X

A List of the Stimuli with Assigned Values and Accepted Responses

Value	Stimulus Word	Correct Responses	Half Correct Responses
1	Full	Empty	Meagre
1	Negative	Affirmative Positive	Assertive
1	After	Before Preceding	Fore
1	Dim	Bright Clear Distinct Luminous	Light Plain Vivid
1	Blunt	Keen Pointed Sharp Tactful	Acute Polite Sensitive Suave

TABLE X—Continued

Value	Stimulus Word	Correct Responses	Half Correct Responses
1	Success	Defeat Disaster Failure	Downfall Loss
1	Pessimistic	Optimistic	Joyful
1	Joy	Gloom Grief Misery Sadness Sorrow	Depression Despondency Displeasure Pain Woe
2	Public	Private	Domestic Personal Secluded Secret
2	Profit	Loss	Deficit Lose
2	Spend	Earn Hoard Husband Keep Save	Accumulate Get Hold Make Receive
2	Always	Never	Infrequently Seldom
2	Graceful	Awkward Clumsy Gawky Ungainly	Uncouth Ungraceful Unsightly
2	Strength	Feebleness Frailty Weakness	Delicacy Insecurity
3	Ancient	Modern	New Present Recent Young
3	Expand	Contract Decrease Diminish Shrink Shrivel	Compress Narrow
3	Barbarous	Civilized Humane	Chivalrous Civil Cultured Educated Gentle Kind Polite

TABLE X—Continued

Value	Stimulus Word	Correct Responses	Half Correct Responses
3	Hinder	Aid Assist Expedite Fore Forward Further Facilitate Help Promote	Advance Encourage
3	Despondent	Buoyant Cheerful Ecstatic Elated Exuberant Glad Happy Hopeful Joyful Jubilant	Bright Encouraged Exhilarated Gay Hilarious Joyous Merry Optimistic Sanguine
3	Vague	Clear Definite Distinct Exact Plain Specific	Apparent Concise Explicit
3	Fertile	Arid Barren Poor Sterile Unproductive	Fruitless Impotent Impoverished Infertile Unimaginative
4	Doubtful	Assured Certain Evident Hopeful Sure	Apparent Believable Clear Confident Credible Known Positive True Truthful Unquestioned
4	Injurious	Advantageous Beneficial Helpful Innocuous Wholesome	Harmless Healthful Safe Uninjurious
4	Busy	Idle Unemployed Unoccupied	Dull Inactive Indolent Loafing

TABLE X—Continued

Value	Stimulus Word	Correct Responses	Half Correct Responses
4	Abstract	Concrete	Specific
4	Advance	Decline Recede Retard Retire Retreat Retrograde Withdraw	Degrade Hinder Withhold
4	Foreign	Domestic Germane Native Pertinent	American Home Indigenous Local Natural
4	Create	Annihilate Demolish Destroy	Abolish Disintegrate Dismember Exterminate Obliterate Undo Waste
4	Simple	Complex Complicated Compound Elaborate Intricate Ornate Wise	Bright Confusing Difficult Gorgeous Grand Hard Learned Sensible Smart
4	Extravagant	Economical Frugal Miserly Niggardly Parsimonious Saving	Careful Close Conservative Economizing Penurious Stingy Thrifty
4	Aristocratic	Bourgeois Common Democratic Plebeian	Humble Low Lowly Ordinary
4	Rare	Abundant Common Dense Frequent Numerous Plentiful Plenteous Usual	Commonplace Often Ordinary Profuse Regular Occasional
4	Dangerous	Harmless Safe	Peaceful

TABLE X—Continued

Value	Stimulus Word	Correct Responses	Half Correct Responses
5	Slovenly	Neat Neatly Tidy Tidily	Careful Carefully Gracefully Precise Trim
5	Defective	Complete Faultless Normal Perfect Sound Whole	Correct Effective Good Healthy
5	Stingy	Generous Lavish Liberal Prodigal	Bounteous Bountiful Extravagant Freehearted Magnanimous Unselfish Unsharing Wasteful
5	Reveal	Conceal Hide Obscure	Cover Secrete Suppress Withhold
5	Diligent	Dilatory Indolent Lazy Slothful	Careless Negligent Inattentive Idle Shiftless
5	Join	Abandon Disjoin Disconnect Dismember Divide Part Separate Sever Sunder	Detach Disassociate Disband Leave Resign Tear Undo Untie Withdraw
6	Impoverish	Enrich Replenish	Aggrandize Fertilize Nourish Strengthen
7	Permanent	Ephemeral Evanescent Fleeting Passing Temporary Transient Transitory	Ephemerate Fluctuating Impermanent Perishable Shifting Transitional Unsubstantial

TABLE X—Continued

Value	Stimulus Word	Correct Responses	Half Correct Responses
7	Elation	Dejection Depression Despair Despondence Despondency	Grief Sadness Shame Sorrow
7	Sinful	Blameless Righteous Sinless	Godly Holy Innocent Just Religious Perfect Pious Pure Upright Virtuous
7	Obnoxious	Agreeable Beneficial Inoffensive Pleasing	Acceptable Attractive Congenial Desirable Harmless Helpful Likable Pleasant Winning
7	Conservative	Extreme Liberal Progressive Radical	Aggressive Extravagant Extremist Wasteful
7	Victorious	Conquered Defeated Vanquished	Beaten Unsuccessful Unvictorious Whipped
7	Obscure	Clear Conspicuous Eminent Lucid Plain Prominent Renowned Reveal	Apparent Bright Disclose Distinct Evident Famous Noticeable Notorious Obvious Unambiguous
8	Proficient	Deficient Inapt Unskilled Unskillful	Backward Incapable Incompetent Inefficient Lacking Unprepared

TABLE X—Continued

Value	Stimulus Word	Correct Responses	Half Correct Responses
8	Rigid	Elastic Flexible Lax Lenient Limber Limp Plastic Relaxed Supple Yielding	Easy Flabby Flaccid Loose Pliable
8	Repulsion	Attraction Cohesion	Acceptance Agreeableness Admiration Love
9	Imaginary	Real	Actual Prosaic Realistic True
9	Permit	Forbid Prevent Refuse Restrain	Deny Disallow Disapprove Object Prohibit
9	Orthodoxy	Heresy Heterodoxy	Catholicity Liberalism Unconventionality Unsoundness
10	Analytical	Synthetic Synthetical	Unanalytical
10	Extrinsic	Intrinsic	Essential Internal
11	Sacred	Defiled Profane Secular Unholy	Common Desecrated Sacrilegious Unconsecrated Ungodly Vulgar
11	Dynamic	Inert Static	Potential Powerless Weak
11	Loquacious	Laconic Reticent Taciturn Silent	Mute Reserved Quiet Untalkative
12	Heterogeneous	Homogeneous	Alike Similar

TABLE X—Continued

Value	Stimulus Word	Correct Responses	Half Correct Responses
12	Spurious	Authentic Genuine Real	Actual Legitimate Pure Sound True Truthful Unadulterated
13	Disastrous	Advantageous Beneficial Helpful Safe	Favorable Fortunate Gainful Harmless Lucky Prosperous
14	Facility	Difficulty	Awkwardness Clumsiness Disability Effort Inconvenience
14	Pride	Humility Lowliness Meekness	Debasement Degradation Humbleness Humiliation
15	Result	Causation Cause	Beginning Commencement Origin Purpose Start
15	Adroit	Awkward Clumsy Inexpert Maladroit Unskillful	Crude Unskilled
17	Parsimony	Extravagance Generosity Lavishness Liberality Prodigality	Freeheartedness Magnanimity Wastefulness
22	Suave	Brusque Impolite Tactless Unpolished	Abrupt Blunt Crude Gruff Harsh Rough Rude Uncouth Undiplomatic Unpleasant Untactful
26	Exoteric	Esoteric	Private

VI.

The Computation of the Relative Difficulty of the Stimuli

As stated previously, when the test was administered, each response given by the subject as well as the reaction time of that response was recorded. On the basis of these records the following data were tabulated for each stimulus: the different responses, the reaction time of each response, and the failures to respond. Records were so kept that information could be obtained not only regarding the number of subjects giving a certain response but also regarding the reaction times of that response. For instance, instead of merely indicating the number of people who gave "young" as an opposite to "ancient," the records show that "young" was given one time in six-fifths of a second, another time in nine-fifths, another in seven, etc.

Furthermore, in order to ascertain when a sufficiently large number of subjects had been tested for the results to be reliable and to be able to indicate the extent of this reliability, as determined by the change brought about by additional sampling, separate records were kept of all tests made prior to the fall of 1918. These will be designated as Group I. Group II consists of the fifty-six records secured in the early fall of 1918, and Group III of the remaining fifty-six records. Since some of the stimuli were added to the list at different times the number of records of subjects in Group I varies from zero to seventy-nine.

After the acceptable responses had been determined these record sheets were scored and the number of correct responses the number of half correct responses, and the number of failures, as well as the total of all three, were tabulated. The number of failures includes both the failures to give any opposite and the failures to give the correct opposite. These results are to be found in Table XI, columns III, IV, V and VI. Referring to this table, we read that for the word "full," there were in Group

1, sixty correct responses, no half correct responses, and three failures, making a total of sixty-three records for that word obtained before the fall of 1918. Reading further in Group II for the same word, we have fifty-four correct and one half correct responses with one failure, thus totaling fifty-six. In Group III we find the following distribution: fifty-five correct and one half-correct.

In computing the percentage of failures, to be found in column VII, two half-correct responses were considered equivalent to one failure. The formula will make clear the procedure followed.

$$\text{Percentage of failures} = \frac{F + \frac{H}{2}}{N} \text{ in which}$$

F = the number of failures,
H = the number of half-correct responses,
N = the total number tested.

Applying this formula to the word "full" we have in Group I,

$$\frac{3 + \frac{0}{2}}{63} = .047$$

Group II does not represent the percentage of failures in Group II alone but a combination of Groups I and II. Applying the formula again for the same word we have $\frac{4 + .5}{119} = .037$ in which 4 equals the sum of the failures in groups I and II, .5 equals half of the half-correct responses in groups I and II, and 119 is the sum of the total number tested. Similarly Group III represents the sum of groups I, II, and III.

This procedure of estimating the three different percentages of failures, instead of only one was adopted because it indicated the degree of reliability of the measure. Very radical changes obviously would mean little reliability. Efforts were made to have our sampling as representative of all the students as possible. Therefore the number tested include graduate students as well as members of the four different classes of undergraduates.

TABLE XI

Data from Which the Percentage of Failures for Each Stimulus was Computed

I Stimulus	II Group	III Number of Correct Responses	IV Number of Half-Correct Responses	V Number of Failures	VI Total Number Tested	VII Percentage of Failures
Full	I	60	0	3	63	.047
	II	54	1	1	56	.037
	III	55	1	0	56	.028
Negative	I	48	0	2	50	.040
	II	53	0	3	56	.047
	III	56	0	0	56	.030
After	I	71	0	3	74	.040
	II	55	0	1	56	.030
	III	53	0	3	56	.037
Dim	I	66	8	1	75	.066
	II	50	3	3	56	.072
	III	42	14	0	56	.089
Blunt	I	60	0	3	63	.947
	II	48	2	6	56	.084
	III	50	0	6	56	.091
Success	I	59	0	4	63	.063
	II	51	1	4	56	.071
	III	44	3	9	56	.108
Pessimistic	I	47	0	2	49	.040
	II	49	1	6	56	.080
	III	47	0	9	56	.108
Joy	I	65	0	3	68	.044
	II	49	3	4	56	.068
	III	48	2	6	56	.086
Public	I	63	0	5	68	.073
	II	46	3	7	56	.108
	III	49	3	4	56	.105
Profit	I	55	7	5	67	.126
	II	52	0	4	56	.101
	III	41	5	10	56	.139
Spend	I	46	9	8	63	.198
	II	51	1	4	56	.142
	III	51	1	4	56	.122
Always	I	52	1	10	63	.166
	II	44	0	12	56	.188
	III	43	2	11	56	.197

TABLE XI—Continued

I Stimulus	II Group	III Number of Correct Responses	IV Number of Half-Correct Responses	V Number of Failures	VI Total Number Tested	VII Percentage of Failures
Graceful	I	50	13	5	68	.169
	II	44	7	5	56	.161
	III	45	4	7	56	.161
Strength	I	56	0	8	64	.125
	II	43	0	13	56	.175
	III	41	0	15	56	.204
Ancient	I	33	26	3	62	.258
	II	32	22	2	56	.245
	III	40	14	2	56	.218
Expand	I	45	1	11	57	.201
	II	43	2	11	56	.207
	III	41	0	15	56	.227
Barbarous	I	57	11	0	68	.080
	II	34	17	5	56	.153
	III	34	12	10	56	.194
Hinder	I	52	3	13	68	.213
	II	44	0	12	56	.213
	III	39	5	12	56	.227
Despondent	I	48	12	3	63	.142
	II	37	15	4	56	.172
	III	39	8	9	56	.191
Vague	I	55	2	10	67	.164
	II	42	2	12	56	.195
	III	38	1	17	56	.231
Fertile	I	54	1	15	70	.221
	II	40	0	16	56	.250
	III	47	0	9	56	.222
Doubtful	I	38	9	10	57	.254
	II	39	9	8	56	.238
	III	38	10	8	56	.236
Injurious	I	48	10	10	68	.220
	II	39	8	9	56	.225
	III	36	6	14	56	.249
Busy	I	39	4	24	67	.388
	II	35	1	20	56	.378
	III	41	1	14	56	.340
Abstract	I	0	0	0	0	.000
	II	35	2	19	56	.357
	III	33	4	19	56	.365

TABLE XI—Continued

I Stimulus	II Group	III Number of Correct Responses	IV Number of Half-Correct Responses	V Number of Failures	VI Total Number Tested	VII Percentage of Failures
Advance	I	48	0	10	58	.172
	II	32	4	20	56	.280
	III	34	1	21	56	.314
Foreign	I	37	10	6	53	.207
	II	34	10	12	56	.256
	III	34	10	12	56	.272
Create	I	50	3	14	67	.231
	II	38	1	17	56	.292
	III	35	2	19	56	.296
Simple	I	13	21	9	43	.453
	II	34	15	7	56	.343
	III	32	14	10	56	.328
Extravagant	I	41	8	7	56	.196
	II	42	6	8	56	.196
	III	35	9	12	56	.229
Aristocratic	I	52	4	12	68	.205
	II	34	5	17	56	.270
	III	32	1	23	56	.316
Rare	I	37	16	8	61	.262
	II	38	7	11	56	.260
	III	38	6	12	56	.262
Dangerous	I	35	1	15	51	.303
	II	36	0	20	56	.331
	III	35	0	21	56	.346
Slovenly	I	44	10	14	68	.279
	II	27	11	18	56	.342
	III	28	4	24	56	.380
Defective	I	47	6	15	68	.264
	II	38	6	12	56	.264
	III	27	8	21	56	322
Stingy	I	43	11	14	68	.286
	II	30	11	15	56	.322
	III	18	13	25	56	.397
Reveal	I	38	4	21	63	.365
	II	31	4	20	56	.378
	III	34	4	18	56	.371
Diligent	I	43	4	20	67	.328
	II	33	15	8	56	.304
	III	26	9	21	56	.351

TABLE XI—Continued

I Stimulus	II Group	III Number of Correct Responses	IV Number of Half-Correct Responses	V Number of Failures	VI Total Number Tested	VII Percentage of Failures
Join	I	33	11	11	55	.300
	II	38	0	18	56	.310
	III	30	7	19	56	.341
Impoverish	I	38	4	27	69	.420
	II	22	3	31	56	.492
	III	27	0	29	56	.500
Permanent	I	37	5	21	63	.373
	II	24	1	31	56	.462
	III	34	1	21	56	.437
Elation	I	33	10	19	62	.387
	II	17	3	36	56	.521
	III	18	3	35	56	.563
Sinful	I	20	40	8	68	.411
	II	13	25	18	56	.471
	III	17	19	20	56	.488
Obnoxious	I	43	11	13	67	.276
	II	14	22	20	56	.402
	III	16	14	26	56	.460
Conservative	I	39	3	27	69	.413
	II	15	15	26	56	.496
	III	14	15	27	56	.533
Victorious	I	43	4	24	67	.388
	II	22	9	25	56	.451
	III	24	8	24	56	.466
Obscure	I	27	4	26	57	.491
	II	25	3	28	56	.508
	III	25	6	25	56	.505
Proficient	I	18	30	15	63	.476
	II	16	18	22	56	.512
	III	14	13	29	56	.551
Rigid	I	27	14	16	57	.403
	II	17	18	21	56	.469
	III	14	18	24	56	.508
Repulsion	I	29	4	34	67	.537
	II	20	2	34	56	.577
	III	18	3	35	56	.600
Imaginary	I	22	4	28	54	.555
	II	21	2	33	56	.581
	III	20	2	34	56	.596

TABLE XI—Continued

I Stimulus	II Group	III Number of Correct Responses	IV Number of Half-Correct Responses	V Number of Failures	VI Total Number Tested	VII Percentage of Failures
Permit	I	28	12	28	68	.500
	II	24	7	25	56	.504
	III	24	8	24	56	.502
Orthodoxy	I	15	0	40	55	.727
	II	9	0	47	56	.783
	III	6	0	50	56	.820
Analytical	I	29	1	27	57	.482
	II	10	1	45	56	.646
	III	11	0	45	56	.698
Extrinsic	I	1	0	1	2	.500
	II	16	0	40	56	.706
	III	15	0	41	56	.719
Sacred	I	24	4	30	58	.551
	II	15	7	34	56	.609
	III	10	11	35	56	.647
Dynamic	I	5	4	30	39	.820
	II	3	4	49	56	.873
	III	7	2	47	56	.867
Loquacious	I	14	20	28	62	.612
	II	13	3	40	56	.673
	III	13	5	38	56	.689
Heterogeneous	I	2	0	1	3	.333
	II	9	4	43	56	.779
	III	12	0	44	56	.782
Spurious	I	6	12	31	49	.755
	II	4	3	49	56	.833
	III	4	3	49	56	.857
Disastrous	I	18	2	30	50	.620
	II	15	6	35	56	.650
	III	21	2	33	56	.635
Facility	I	10	1	34	45	.766
	II	6	3	47	56	.821
	III	5	0	51	56	.853
Pride	I	17	3	53	73	.746
	II	8	2	46	56	.786
	III	6	1	49	56	.816
Result	I	15	15	32	62	.636
	II	10	6	40	56	.698
	III	4	6	46	56	.755

TABLE XI—Continued

I Stimulus	II Group	III Number of Correct Responses	IV Number of Half-Correct Responses	V Number of Failures	VI Total Number Tested	VII Percentage of Failures
Adroit	I	19	2	40	61	.672
	II	7	3	46	56	.756
	III	3	0	53	56	.817
Parsimony	I	11	2	35	48	.750
	II	7	0	49	56	.816
	III	6	6	44	56	.823
Suave	I	2	10	34	46	.847
	II	2	9	45	56	.867
	III	1	8	47	56	.882
Exoteric	I	3	0	49	52	.942
	II	1	0	55	56	.962
	III	0	0	56	56	.975

Since the value of the stimuli was to be in terms of both accuracy and speed, the median time of all responses, both correct and half correct, was computed. The time score taken with a stop-watch represents fifths of a second. As in the case of the accuracy score, three time scores were reckoned for each word, the last score representing the combined time of all three groups. See column III, Table XII. The percent of failures of the preceding table is copied in order to make clear the process by which the values of the several stimuli are reached.

To arrive at the relative difficulty of the stimuli, the percentage of failures was multiplied by the median time. If credit is to be given for both measures, some more or less arbitrary scheme must be adopted. This procedure seemed justifiable since by it words of equal accuracy scores received different point values if the difference in time varied to any appreciable extent. This was precisely what was desired. Any method which covered up this difference was to be avoided.

In order to determine a point value for the stimuli, the standard deviation of the different values, as indicated in the fourth column of the following table, was obtained. Beginning at the average and calling it zero, steps of one-fifth standard deviation were then made, positively and negatively. These steps being

equal in a sense, the furtherest negative step was assigned a value of one, the next two, and so on up to twenty-six, the most difficult word in the list. These point values are given in the fifth column of Table XII.

In selecting the stimuli for the completed test, those words which had received a hundred percent accuracy score were discarded, for thus we were better enabled to locate a zero point. Obviously we cannot find or infer the place where the ability to give opposites begins, as we can in the physical sciences for weight or length. Therefore the zero point will of necessity be an arbitrary one selected to represent very little of the trait in question. In eliminating all words with a hundred percent accuracy score, but retaining those next in difficulty, we assume that a college student who is unable to score a point on the test as it now stands has zero ability in naming opposites. Since we have located the distance from the median in terms of standard deviation, we have equally difficult steps. Roughly we assigned to the furtherest negative step a value of one. Since we are not concerned as were Woody[1] and Hering[2] with finding the distances between different grades or groups of pupils, it seems impracticable to be too particular about the absolute point of the zero. In addition any error in slightly misplacing the point might be of less importance than that made in weighting time and accuracy.

It is regretted that, especially in the case of the more difficult words, the measure of time is based on so small a number of records. As an indication of the reliability of the median time, we have computed the median deviation and coefficient of variability for each word. The probable error of the median is calculated by the formula, P. E. med. $= \frac{5\ Q\ \text{dis}}{4\sqrt{n}}$.[3]

[1] Woody, Clifford, "Measurements of Some Achievements in Arithmetic," *Col. Univ. Cont. Educ.*, 1916, LXXX.

[2] Hering, John P., "Derivation of a Scale to Measure Abilities in Scientific Thinking," *Jour. Educ. Psychol.*, 1919, IX, 417-431.

[3] Thorndike, E. L., "Mental and Social Measurements," New York, 1916, p. 195.

TABLE XII

The Point Value of the Stimuli Computed from Table IX

I Stimulus	II Percent of Failures	III Median Time	IV Relative Value	V Point Value	VI Median Deviation	VII Coefficient of Variability	VIII Probable Error of Median
Full	.047	7	.329				
	.037	6	.222				
	.028	7	.209	1	1.48	.21	.14
Negative	.040	7	.280				
	.047	7	.329				
	.030	7	.210	1	1.22	.17	.12
After	.040	6	.240				
	.030	7	.210				
	.037	7	.259	1	1.20	.17	.11
Dim	.066	8	.528				
	.072	7	.504				
	.089	7	.623	1	1.64	.23	.15
Blunt	.047	6	.282				
	.084	7	.588				
	.091	7	.637	1	1.40	.20	.13
Success	.063	6	.378				
	.071	7	.497				
	.108	6	.648	1	1.50	.25	.14
Pessimistic	.040	6	.240				
	.080	6	.480				
	.108	6	.648	1	1.11	.18	.11
Joy	.044	7	.308				
	.068	8	.544				
	.086	8	.688	1	2.08	.26	.20
Public	.073	7	.511				
	.108	7	.756				
	.105	8	.840	2	2.03	.25	.19
Profit	.126	7	.882				
	.101	8	.808				
	.139	8	1.112	2	2.26	.28	.22
Spend	.198	10	1.980				
	.142	10	1.420				
	.122	10	1.220	2	2.80	.28	.27
Always	.166	8	1.328				
	.188	8	1.504				
	.197	8	1.576	2	1.90	.23	.19
Graceful	.169	10	1.690				
	.161	10	1.610				
	.161	10	1.610	2	3.36	.33	.32

TABLE XII—Continued

I Stimulus	II Percent of Failures	III Median Time	IV Relative Value	V Point Value	VI Median Deviation	VII Coefficient of Variability	VIII Probable Error of Median
Strength	.125	7	.875				
	.175	8	1.400				
	.204	8	1.632	2	1.95	.24	.20
Ancient	.258	8	2.064				
	.245	7.5	1.837				
	.218	8	1.744	3	2.30	.25	.22
Expand	.201	9	1.809				
	.297	8	1.656				
	.227	8	1.816	3	1.28	.16	.13
Barbarous	.080	11	.880				
	.153	11	1.683				
	.194	11	2.134	3	2.90	.26	.28
Hinder	.213	9	1.917				
	.213	10	2.130				
	.227	10	2.270	3	3.48	.34	.36
Despondent	.142	14	1.988				
	.172	12	2.064				
	.191	12	2.292	3	3.54	.29	.35
Vague	.164	9	1.476				
	.195	10	1.950				
	.231	11	2.541	3	3.23	.29	.34
Fertile	.221	12	2.652				
	.250	12	3.000				
	.222	12	2.664	3	3.85	.32	.40
Doubtful	.254	12	3.048				
	.238	12	2.858				
	.236	12	2.832	4	3.89	.32	.40
Injurious	.220	11	2.420				
	.225	12	2.700				
	.249	12	2.988	4	4.47	.37	.46
Busy	.388	10	3.880				
	.378	10	3.780				
	.340	9	3.060	4	2.64	.29	.30
Abstract	.000	0	0.000				
	.357	9	3.213				
	.365	8.5	3.102	4	2.10	.24	.30
Advance	.172	9	1.548				
	.280	9	2.520				
	.314	10	3.140	4	2.90	.29	.33

TABLE XII—Continued

I Stimulus	II Percent of Failures	III Median Time	IV Relative Value	V Point Value	VI Median Deviation	VII Coefficient of Variability	VIII Probable Error of Median
Foreign	.207	12	2.484				
	.256	12	3.072				
	.272	12	3.264	4	3.80	.31	.45
Create	.231	10	2.310				
	.292	11	3.212				
	.296	11	3.256	4	3.31	.30	.36
Simple	.453	8	3.624				
	.343	10	3.430				
	.328	10	3.280	4	3.23	.32	.28
Extravagant	.196	14	2.744				
	.196	15	2.940				
	.229	15	3.435	4	4.14	.27	.43
Aristocratic	.205	12	2.460				
	.270	11	2.970				
	.316	11	3.476	4	4.20	.38	.46
Rare	.262	14	3.668				
	.260	14	3.640				
	.262	14	3.668	4	3.39	.24	.35
Dangerous	.303	11	3.333				
	.331	11	3.641				
	.346	11	3.806	4	3.75	.34	.45
Slovenly	.279	11	3.069				
	.342	12	4.104				
	.380	11	4.180	5	3.62	.32	.63
Defective	.264	13	3.432				
	.264	13	3.432				
	.322	13	4.186	5	5.29	.40	.58
Stingy	.286	10	2.860				
	.322	12	3.864				
	.397	11	4.367	5	3.34	.30	.37
Reveal	.365	11	4.015				
	.378	11	4.158				
	.371	12	4.452	5	3.83	.31	.44
Diligent	.328	13	4.264				
	.304	13	3.952				
	.351	13	4.563	5	3.97	.30	.43
Join	.300	14	4.200				
	.310	13	4.030				
	.341	14	4.774	5	4.90	.35	.56

TABLE XII—Continued

I Stimulus	II Percent of Failures	III Median Time	IV Relative Value	V Point Value	VI Median Deviation	VII Coefficient of Variability	VIII Probable Error of Median
Impoverish	.420	12	5.040				
	.492	11	5.412				
	.500	11	5.500	6	.341	.31	.49
Permanent	.373	13	4.849				
	,462	13	6.006				
	.437	14	6.118	7	4.50	.32	.55
Elation	.387	10	3.870				
	.521	12	6.262				
	.563	12	6.756	7	3.79	.31	.51
Sinful	.411	13	5.343				
	.471	16	7.536				
	.488	14	6.832	7	5.10	.36	.64
Obnoxious	.276	13	3.588				
	.402	14	5.628				
	.460	15	6.900	7	5.36	.35	.61
Conservative	.413	11	4.543				
	.496	11	6.448				
	.533	13	6.929	7	4.30	.33	.53
Victorious	.388	14	5.432				
	.451	15	6.765				
	.466	15	6.990	7	4.75	.31	.56
Obscure	.491	12	5.892				
	.508	13	6.604				
	.505	14	7.070	7	6.06	.43	.79
Proficient	.476	11	5.236				
	.512	13	6.656				
	.551	13	7.163	8	5.45	.41	.65
Rigid	.403	15	6.045				
	.469	15	7.035				
	.508	15	7.620	8	4.97	.33	.59
Repulsion	.537	11	5.907				
	.577	13	7.501				
	.600	13	7.800	8	5.30	.40	.75
Imaginary	.555	11	6.105				
	.581	12	6.972				
	.596	14	8.344	9	4.56	.32	.67
Permit	.500	16	8.000				
	.504	16	8.064				
	.502	17	8.534	9	6.50	.38	.80

TABLE XII—Continued

I Stimulus	II Percent of Failures	III Median Time	IV Relative Value	V Point Value	VI Median Deviation	VII Coefficient of Variability	VIII Probable Error of Median
Orthodoxy	.727	11	7.997				
	.783	11	8.613				
	.820	11	9.020	9	2.50	.22	.57
Analytical	.482	16.5	7.953				
	.646	14	9.044				
	.698	14	9.772	10	4.27	.30	.74
Extrinsic	.500	9	4.500				
	.706	17	12.000				
	.719	14	10.066	10	5.30	.37	1.17
Sacred	.551	15	8.265				
	.609	16	9.744				
	.647	16	10.352	11	6.17	.38	.91
Dynamic	.820	12	9.840				
	.873	12	10.476				
	.867	12	10.404	11	3.75	.31	.93
Loquacious	.612	15	9.180				
	.673	15	10.095				
	.689	16	11.024	11	5.08	.31	.76
Heterogeneous	.333	14.5	4.828				
	.779	13	10.127				
	.782	15	11.730	12	4.88	.32	1.17
Spurious	.755	14	10.570				
	.833	14	11.662				
	.857	14	11.998	12	5.50	.39	1.21
Disastrous	.620	17	10.540				
	.650	19	12.350				
	.635	20	12.700	13	7.08	.35	1.10
Facility	.766	16	12.256				
	.821	18	14.778				
	.853	16	13.648	14	7.00	.43	1.75
Pride	.746	16	11.936				
	.786	17	13.362				
	.816	17	13.872	14	5.88	.34	1.20
Result	.636	20	12.740				
	.699	22	15.378				
	.755	20	15.100	15	8.00	.40	1.33
Adroit	.672	16	10.752				
	.756	17	12.852				
	.817	18.5	15.114	15	6.23	.33	1.33

TABLE XII—Continued

I Stimulus	II Percent of Failures	III Median Time	IV Relative Value	V Point Value	VI Median Deviation	VII Coefficient of Variability	VIII Probable Error of Median
Parsimony	.750	25	18.750				
	.816	22	17.952				
	.823	21	17.183	17	6.60	.31	1.45
Suave	.847	24.5	20.751				
	.867	25	21.675				
	.882	25	22.050	22	9.17	.36	2.02
Exoteric	.942	36	33.912				
	.962	27.5	26.455				
	.975	27.5	26.812	26	9.50	.34	5.93

VII.

The Conversion into a Group Test

In order to make it possible for even the inexperienced to use the test with facility, it was converted into a group test. Thus, obviously, much time in administering and labor in scoring were dispensed with. In addition, the personal element became almost negligible. The words were arranged in the ascending order of difficulty and six minutes allowed for the test. This time limit was determined empirically, after experimenting upon a dozen or more students at the George Peabody College and some sixty students at the North Georgia Agricultural College.[1] In arriving at the time to be allowed, two considerations were held in mind. It was desired that sufficient time be given for individual scores to cover a considerable range, and thus avoid a large undistributed group of poor records. On the other hand the time should be so planned that a perfect score is impossible. Two individuals with a perfect score might vary in efficiency, but under the conditions of the test no measure of their difference could be obtained.

The test sheets were printed upon a good quality of white paper, eight and a half by eleven inches. The directions printed on the back of the sheet explained to the experimenter as well as to the subject, exactly what was to be done. A copy is included in the Appendix. By mistake the word "rare" was placed between "defective" and stingy." It should follow "aristocratic." In scoring the papers, however, proper credit was assigned the word. The displacement was so slight that it is highly improbable that the final results were at all affected.

Three thousand copies of the test were sent to psychologists in twenty different normal schools, colleges, and universities throughout the country. Of this number, one thousand six hundred and twenty-eight copies were returned for grading from

[1] These results, which were obtained through the kindness of Professor George Camp, are not included in the norms because the sheets were mimeographed instead of printed.

fourteen widely different institutions. Unfortunately, a number of papers arrived too late to be included in the norms herein presented. The table below indicates the number of subjects tested in each school and the class to which both the men and women then belonged.

TABLE XIII

The Number of Men and Women Tested in Each School and Class in Which They Were Enrolled

	Freshmen		Soph's		Juniors		Seniors		Graduates		Total
	Women	Men	Women	Men	Women	Men	Women	Men	Women	Men	
Wellesley College			43		95						138
Brooklyn Training School	148	5									153
University of Illinois	4	2	44	46	60	34	35	16		2	243
University of Montana			23	2	10	5	24	1			66
Alabama State Normal			50								50
University of Chicago	2	6	40	74	7	13	20	9	2		173
Teachers' College					2		13	2	9	28	54
Vanderbilt University						1			1	12	14
University of Kansas		1	32	27	11	6	3	4	1		85
University of Tennessee	1		25	19	7	2	4	2			60
University of Louisiana	2	1	14	9	5	14	8	4	1		58
Fitchburg Normal School	54	17	26		17	2					116
Kansas State Normal	49	6	38	1	15	3					112
Ohio State University	59	63	40	58	33	22	14	13	2	2	306
Total	319	101	375	237	262	102	121	51	16	44	1628

The final results of the sixteen hundred records indicated class variations as well as slight sex differences. In order to ascertain a possible cause, the writer has computed the percentage of students of both sexes in each of the classes. This distribution is shown in the following table.

In grading the papers two classes of errors were encountered, namely, the misspelled word, and the abbreviated word. In the former case full credit was allowed whenever the spelling clearly

TABLE XIV

The Percentage of Students of Both Sexes in Each of the Classes

	Freshmen	Sophomores	Juniors	Seniors	Graduates
Women	.291	.343	.239	.110	.014
Men	.188	.440	.190	.095	.082
Both Sexes	.27	.37	.22	.10	.03

showed that an acceptable word was meant. However, when a word similar in sound was correctly spelled no credit was given. As an example, "save" was not accepted as an opposite of "dangerous," but "richous" (actual case) as an opposite of "sinful" was given full credit. This rule was adopted because without it gradings by different individuals would vary considerably.

Where the response was abbreviated it was given no credit. The only two abbreviations were "opt" as the opposite of pessimistic and "dem" as opposite of aristocratic. In both cases the subject might have intended to write "optimist" and "democrat," both of which would have been scored a failure. Even had he had in mind the adjective instead of the noun, his score would have been raised only five points. No doubt in abbreviating the words, he was thus enabled, at an unfair advantage to others, to write the opposite of a more difficult word. In view of the fact that a time limit was placed on the test, not to penalize abbreviations seems unjust to those students who have followed the directions carefully. Fortunately the number of students abbreviating was negligible.

The sum of the point values of the stimuli as indicated in Table XII is four hundred and thirty-eight. Since each response may receive the assigned value, one-half that value, or zero, almost any score between zero and this number is possible. This wide range of distribution eliminates the probability of a large group of subjects of different ability receiving identical scores. The actual scores based on the records obtained varied from zero to three hundred and fifty-one.

VIII.

The Establishment of Norms for College Students

Sixteen hundred and twenty-eight test papers were graded and the norms for college students by the group test were based on these records. In order to ascertain the form of the distribution of the scores made by the several classes, steps of fifty beginning with zero were made and the percent of students included between these steps was estimated. Examination of Table XV reveals the fact that the largest percentage of scores falls between one hundred and one hundred fifty. This is perhaps more clearly shown in the graphs (Figures 2, 3, 4, 5, 6) based on this table.

TABLE XV

Percent of Students of Several Classes Attaining Scores Within the Indicated Limits

Score	Freshmen	Sophomores	Juniors	Seniors	Graduates
0- 49.5	.069	.033	.016	.005	.000
50- 99.5	.356	.272	.190	.184	.066
100-149.5	.397	.346	.318	.406	.300
150-199.5	.124	.225	.247	.214	.250
200-249.5	.045	.078	.156	.111	.163
250-299.5	.007	.032	.063	.047	.133
300-349.5	.002	.014	.010	.028	.086
350-399.5	.000	.000	.000	.005	.000

The norms for college students are presented in the form of percentile tables because this method, statistically simple, admits of convenient use and permits comparisons to be made with other tests, similarly treated. Let us suppose that a directions test has been standardized and the norms compiled in terms of percentiles. Scores made in the opposites test can be compared with scores made in the directions test when both are converted into percentiles. The two percentiles can be averaged or the median computed. This is statistically impossible when the two scores are given in terms of the score in points made on

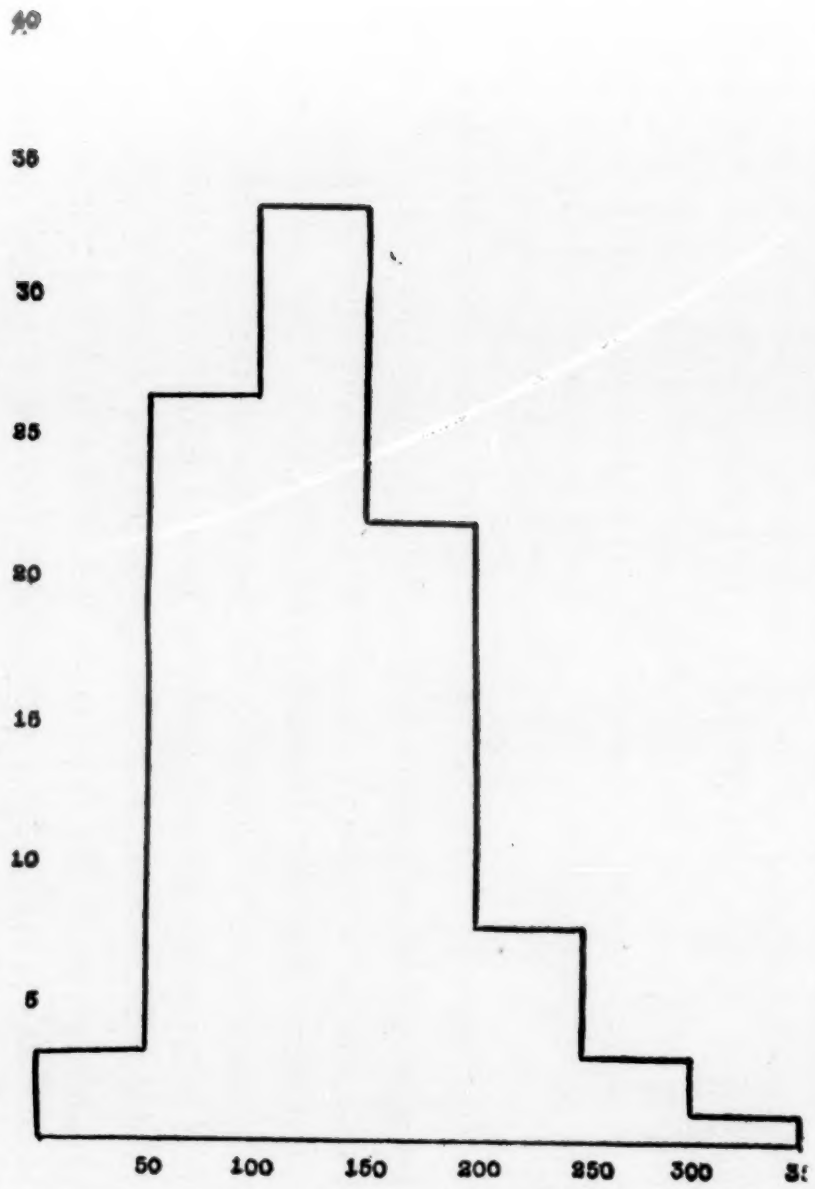

Figure 3
The vertical line shows the percentage of sophomores receiving the scores indicated on the base line.

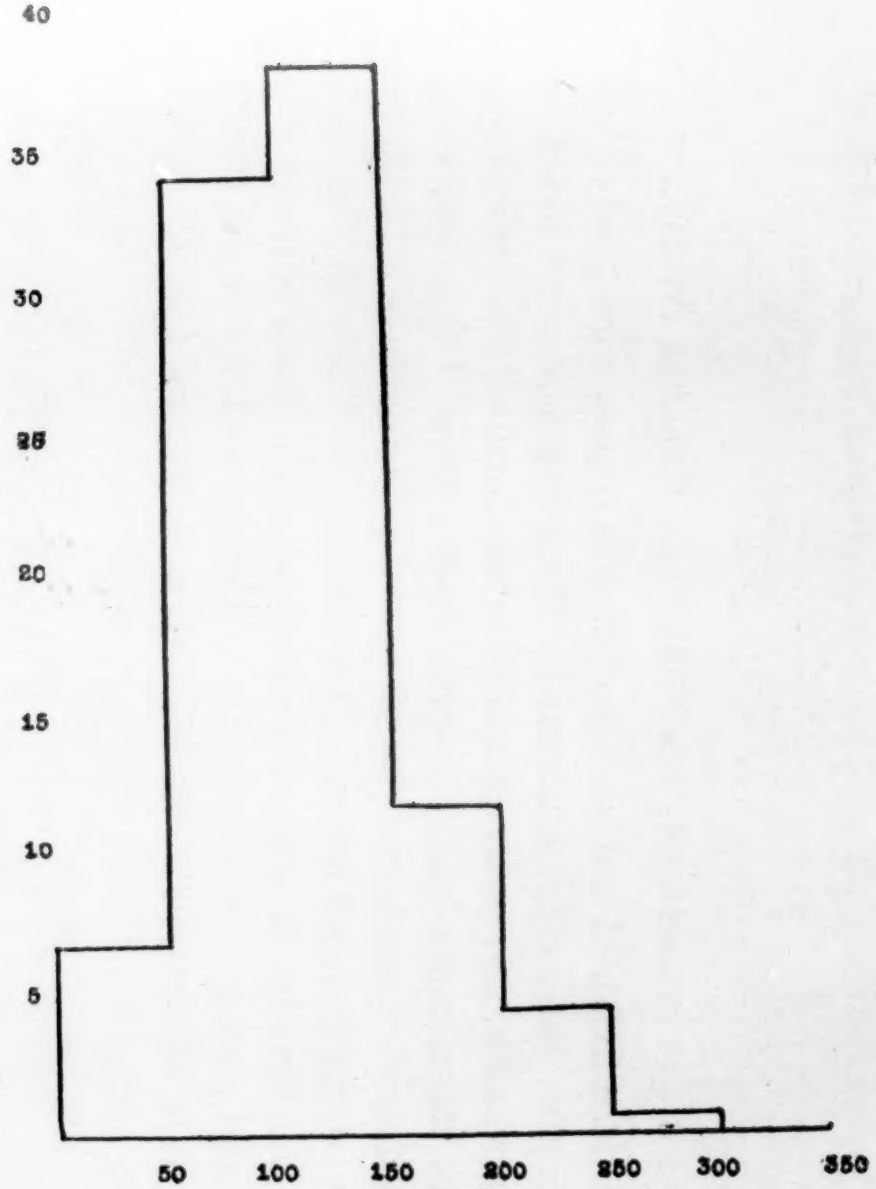

Figure 2
The vertical line of the graph indicates the percentage of freshmen receiving scores shown on the base line.

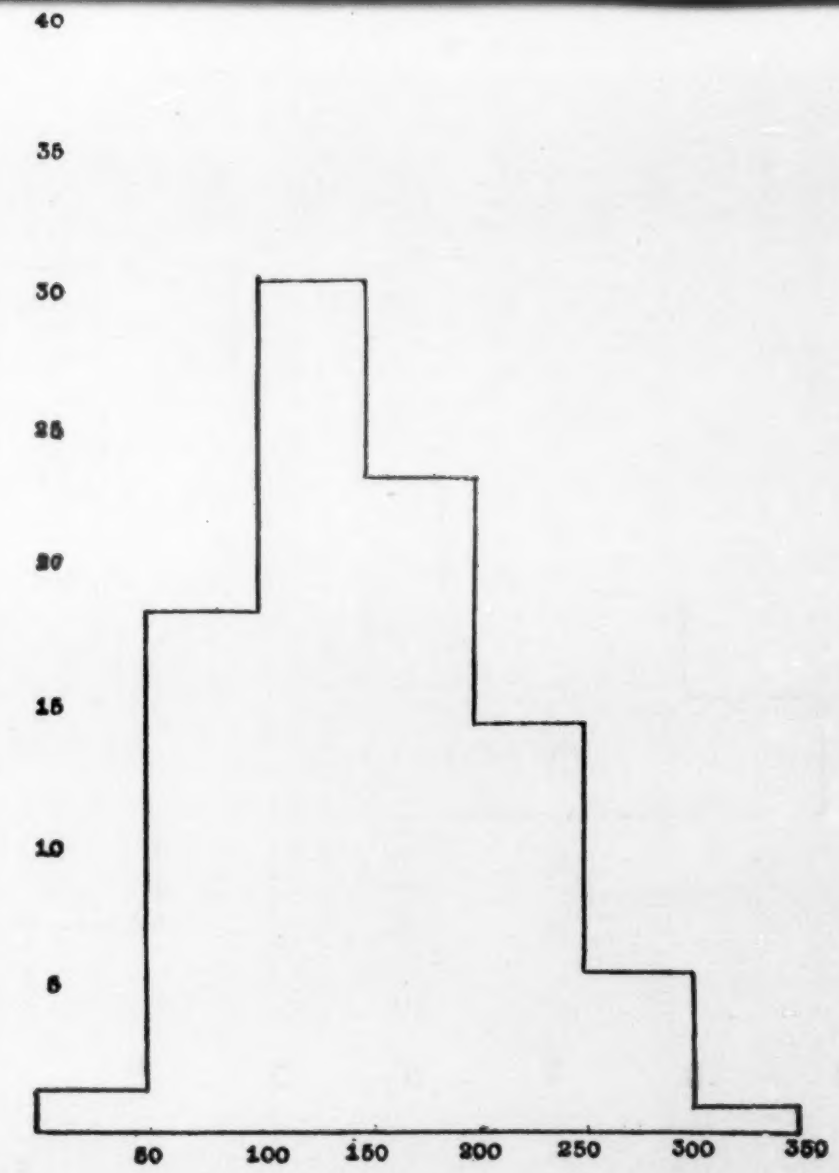

Figure 4
The vertical line of the graph shows the percentage of juniors receiving the scores indicated on the base line.

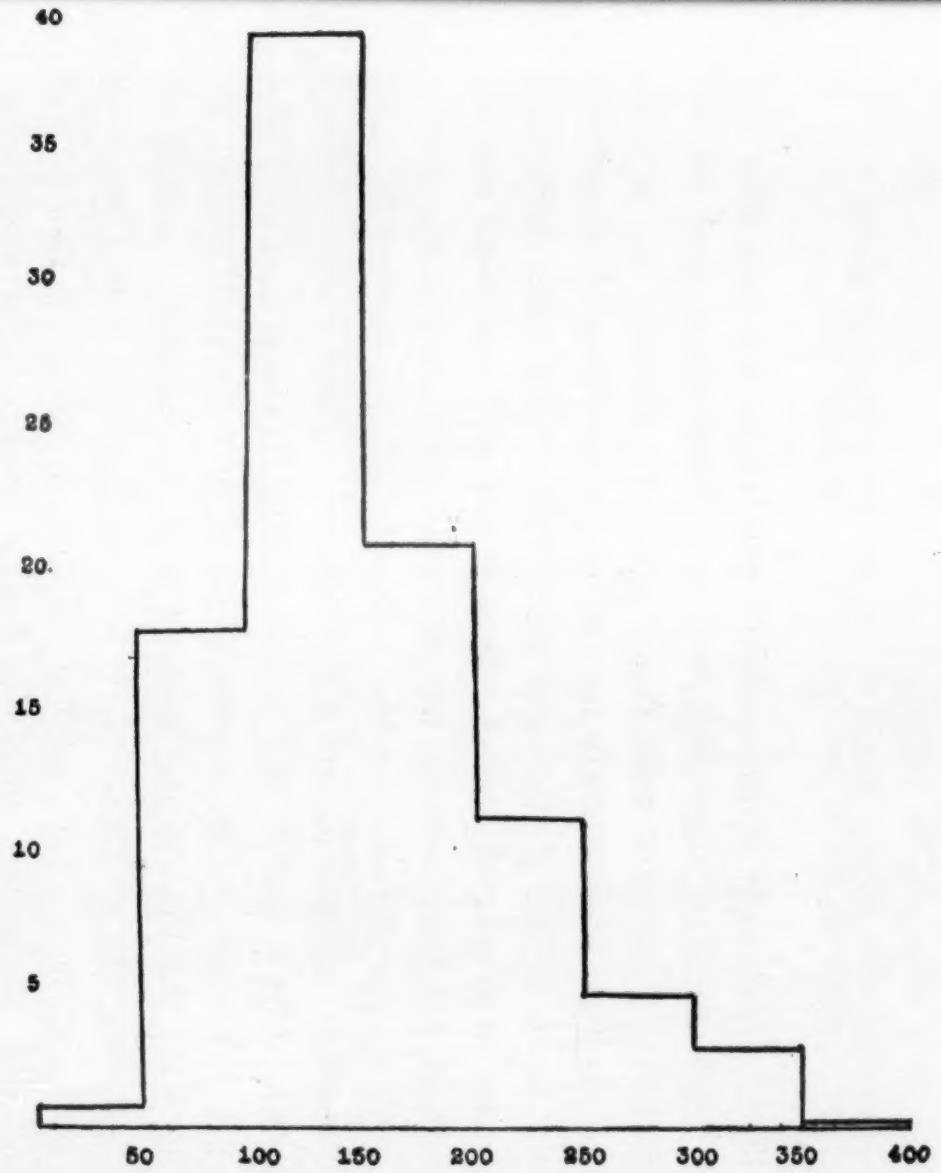

Figure 5
The vertical line in the graph shows the percentage of seniors receiving the scores indicated on the base line.

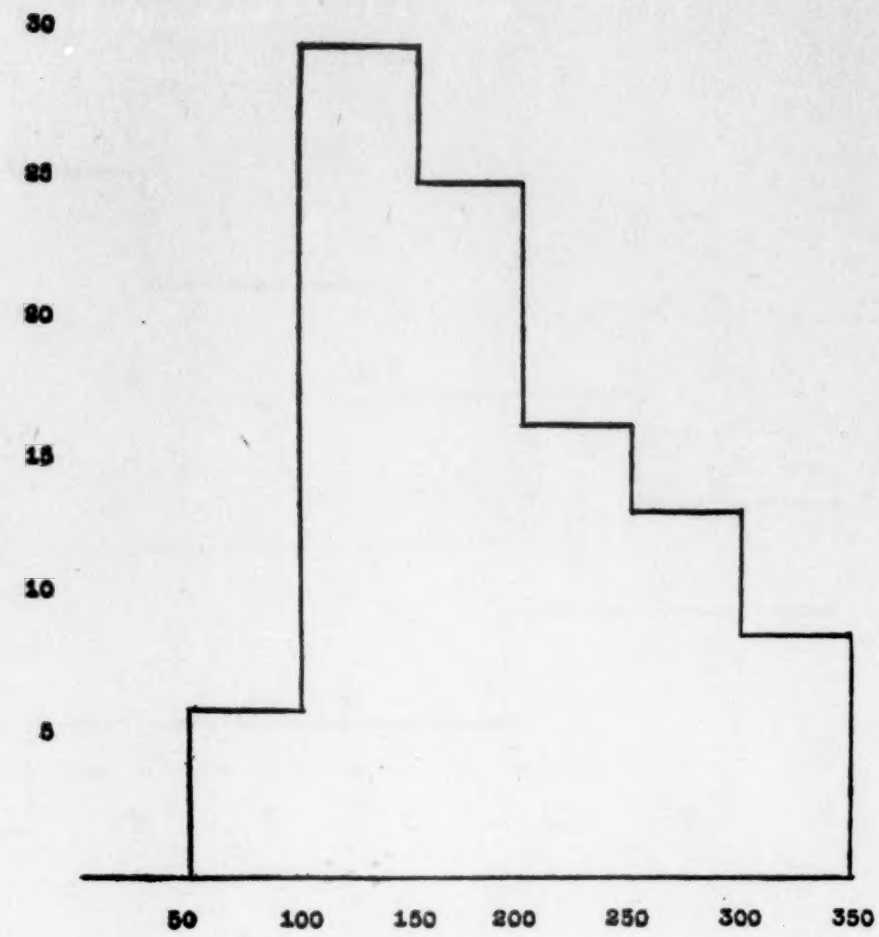

Figure 6

The vertical line in the graph shows the percentage of graduates receiving the scores indicated on the base line.

each test. Obviously, if the possible grades on the opposites test range from zero to four hundred and thirty-eight and on the hypothetical directions test from zero to sixteen, to add or average the two scores made by any individual would practically amount to weighting the opposites test. But the percentile scores can be averaged and the final result will represent a combination of the two tests to which equal credit is given.

In order to make clearer the meaning of the percentile table the following explanation of the manner in which it was derived is offered. For instance in the case of the freshmen, as indicated in Table XVI, all scores were arranged in numerical order, beginning with the lowest and ending with the highest. The number of scores was ascertained, which in this case was four hundred and twenty. Ten percent of the number of scores was counted off, beginning with the lowest score. The score upon which the forty second count fell became the tenth percentile. A

glance at the table will show that in this instance it was fifty-eight and a half. This means that this grade is not exceeded by the lowest ten percent of the freshmen, and that any student making this grade is equalled or surpassed by ninety percent of the college students.

An additional ten percent is counted off to secure the next higher percentile, and so on.

TABLE XVI

Norms for College Students in Terms of Percentile Scores in the Group Test

Percentile	Freshmen	Sophomores	Juniors	Seniors	Graduates	All Comb'd
10	58.5	66.5	77.0	84.0	107.0	68.5
20	74.5	84.0	97.0	100.0	124.0	78.0
30	87.0	99.0	116.0	116.5	138.5	101.0
40	97.0	112.0	130.5	125.5	151.0	114.5
50	107.0	125.0	146.0	137.5	155.5	128.5
60	120.0	139.9	161.5	147.0	191.5	142.0
70	132.5	156.0	181.0	166.0	200.0	157.0
80	147.0	178.5	203.0	187.5	247.0	181.5
90	172.0	214.0	239.0	231.0	298.0	220.5
100	303.5	338.0	313.0	351.0	313.0	351.0
Number Tested	420	612	364	172	60	1628

Frequently comparisons of different groups are stated in terms of the percent of individuals of the first group reaching or exceeding the twenty-five or seventy-five percentile of the second group. In the table below the twenty-five, fifty, and seventy-five percentiles are indicated for each class.

TABLE XVII

Twenty-five, Fifty, and Seventy-five Percentile Scores for Each Class

Percentile	Freshmen	Sophomores	Juniors	Seniors	Graduates	All Comb'd
25	82.5	92.5	106.0	108.0	136.0	89.5
50	107.0	125.0	146.0	137.5	155.5	128.5
75	139.0	165.5	192.0	173.0	213.0	169.0

The sophomores surpass the freshmen and they are in turn outstripped by the juniors. The gap between seniors and graduates is wide; but the former are little if at all superior to the juniors. A consideration of the lowest and highest score made by the several classes reveals the fact that although ten percent of

the juniors barely reach the lowest score made by the graduates, still the highest score attained by each class is identical. In Table XVIII is shown the gradual rise of the lowest score realized by the different classes, contrasted with the irregularity of the highest score.

TABLE XVIII

Table Showing the Highest and Lowest Score Made by Each Class

	Freshmen	Sophomores	Juniors	Seniors	Graduates
Lowest Score	0	18.5	38.0	46.0	75.0
Highest Score	303.5	338.0	313.0	351.0	313.0

Although men are slightly superior in the test, to women belongs the highest as well as the lowest score. The greatest sex difference is to be found in the fortieth and fiftieth percentiles which in each case are surpassed by the next lowest percentile secured by the men. The table below shows the ten percentiles for both sexes and the number tested in each case. It also includes, for the benefit of those interested in sex differences, the more commonly employed twenty-five and seventy-five percentiles.

TABLE XIX

Percentile Scores for Men and Women

Percentiles	Men	Women
0		
10	72.0	60.0
20	92.5	85.5
25	99.0	92.0
30	106.0	98.5
40	122.5	101.0
50	133.5	115.0
60	147.0	139.0
70	163.5	155.5
75	176.0	167.0
80	189.0	178.0
90	221.0	222.0
100	338.0	351.0
Number Tested	535	1093

A consideration of the range of scores made by both sexes would seem to indicate that women have a wider range than men in the trait in question. Whereas men range only from 18.5 to 338, women beginning with an initial score of 0 reach 351 points as their highest score.

IX.

Interpretation of the Results

Whether the test is, as Simpson[1] believes, a selective thinking test, we cannot from the very nature of the data conclude. However, we feel justified in assuming that it is a test success in which is dependent upon native ability rather than number of years schooling. The higher scores attained by the graduates confirm rather than refute the contention, for although there is a constant process of weeding out the mentally inferior, selection operates particularly at the conclusion of work offered for the bachelor's degree. The graphs (figures 2, 3, 4, 5, and 6) reveal a considerable amount of overlapping from year to year. Between ten and twenty percent of the freshmen reach the median score of the graduates and practically thirty percent, the median score of the seniors.

Not without significance is the range of scores of the several classes. An examination of Table XVIII discloses the fact that the highest individual score obtained by either a junior or graduate is surpassed by a sophomore. This condition is not to be attributed to the simplicity of the test, which might admit of a number of nearly perfect records. On the contrary three hundred and fifty-one is the highest score reached, whereas a possible four hundred and thirty-eight could be attained in the allotted time were an individual familiar with the opposites. The lowest score made by the several classes ascends regularly from freshmen to graduates, all of which seems to indicate that instead of the scores progressing in absolute value as the years in school increase, the limit of attainment is about reached in the freshmen year, and thereafter a dropping off of the poorer students raises the percentage of high scores made by the more advanced students.

The form of the distribution of the scores, skewed as it is to

[1] Simpson, Benjamin R., *Op. cit.*

the lower end, indicates that the time allotted is too short. Had eight minutes instead of six been allowed for the test, the scores would have been scattered over a wider area and the form of distribution would probably have approached the normal. As it stands the test provides a large number of possible scores which can be attained by the college student. In addition, because of its difficulty, it offers an opportunity of testing a more highly selected group.

Sex differences favor the male. King and M'Crory[2] found that the same condition obtained with the hard opposites test they used. In the present investigation twice as many women were tested as men. An equal number might alter the situation. This difference might be due to the fact that the normal schools largely attended by women were slightly inferior in the test to the colleges and universities. In addition, proportionally more women belonged to the freshman class and fewer were enrolled in the graduate school, as is indicated in Table XIV. Until an equal number of unselected cases from each sex is tested, the subject is debatable.

[2] King, Irving, and M'Crory, J. L., "Freshmen Tests at the State University of Iowa," *Jour. Educ. Psychol.*, 1918, IX.

X.

Suggestions and Directions for the Use of the Opposites Test

The following suggestions are recommended for using the test: Distribute the test sheets with the face downward and read aloud the written directions found on the back of the sheets. In timing the class, make use of a stop-watch. Allowing the point value of the word for correct responses and half that value for half correct responses, grade the papers according to the acceptable responses presented in Table X. Add the points and record this sum as the individual's score.

If an entire class is to be compared with the norms herein contained, compute the median score of the class and compare with the fifty percentile to ascertain whether the class is above or below standard. The ratio between the percentile score of the individual or of the class and the fifty percentile of the group may be taken as index of brightness.

Frequently it is necessary to have more than one list of opposites. To meet such a contingency, the list has been divided so that both parts contain an equal number of points and words. The lists are given below, arranged in the order of difficulty. The assigned point value is at the left of the stimulus.

The writer feels that the Opposites Test as a group test has sacrificed to time an important element in that it fails to detect certain aspects of the subject's mind, aspects upon which personal efficiency is so largely dependent. Most significant and enlightening is the attitude of the subject toward the test.

LIST I

1 Full	3 Despondent	5 Reveal	10 Analytical
1 Dim	4 Doubtful	5 Join	11 Loquacious
1 Blunt	4 Busy	7 Permanent	11 Sacred
1 Pessimistic	4 Abstract	7 Obnoxious	12 Spurious
2 Public	4 Foreign	7 Victorious	14 Facility
2 Spend	4 Simple	8 Proficient	14 Pride
2 Graceful	4 Extravagant	8 Rigid	17 Parsimony
3 Ancient	4 Dangerous	9 Orthodoxy	22 Suave
3 Barbarous	5 Defective		

LIST II

1 Negative	3 Vague	5 Diligent	9 Permit
1 After	3 Fertile	6 Impoverish	10 Extrinsic
1 Success	4 Injurious	7 Elation	11 Dynamic
1 Joy	4 Advance	7 Sinful	12 Heterogeneous
2 Profit	4 Create	7 Conservative	13 Disastrous
2 Always	4 Aristocratic	7 Obscure	15 Result
2 Strength	4 Rare	8 Repulsion	15 Adroit
3 Expand	5 Slovenly	9 Imaginary	26 Exoteric
3 Hinder	5 Stingy		

Whether these attitudes, which Ruger[1] has entitled subjective and objective, so easily perceived by the experimenter, carry over into other lines of work, while highly probable, is a matter that yet remains to be proved.

The poor record in the Opposites Test does not necessarily indicate the subjective attitude but an exceptionally good record does more probably indicate the absence of it. The student becomes easily confused and his attention is divided between the opposite to be given and fear of a bad showing. The score goes low on account of frequent interruptions of which the following are typical: "I know the opposite but I can't think of it right now," "I can't think of the opposite but I know the word perfectly," "This is a vocabulary test and my vocabulary is limited."

This type of subject seldom admits that he is familiar neither with the word nor the opposite. He prefers in the individual test to waste the allotted ten seconds attempting to create the impression that he knows the opposite but this time he can not quite get it. Frequently he forgets the present stimulus in regretting the one just missed.

However if time is wasted in administering the test, it is more than counterbalanced by the speed with which he makes his exit. He does not tarry to ask about some of the words he has failed on, nor is he curious about the records of others. With a hasty explanation of his own inefficiency he hurries from the room, leaving the experimenter with a vague sort of feeling that the latter has purposely inflicted a mortal injury.

In striking contrast is the subject whose interest is centered upon the test itself. He works rapidly offering neither apologies

[1] Ruger, H. A., "The Psychology of Efficiency, *Archiv. of Psychol.*, 1910, XV.

nor explanations. Unashamed he denies knowledge of some of the more difficult words, but rarely does he depart without this knowledge. He is eager to know what others are able to do under the same conditions. Interest in self is overshadowed by interest in the problem in hand.

Peterson,[2] with his Rational Learning Test, notes these same attitudes, and the writer knows of no better test in which to observe them, for not only are they evident during the administration of the test, but the traits in question can be easily detected in the individual records.

Determination of the extent to which the opposites test reveals native ability, is not attempted in this research. It is hoped that someone will carry on the work here begun, and by a series of correlations empirically determine the scope and limitation of the test. The writer is of the opinion, that if used with other standardized tests, the Opposites Test will be found of value in mental diagnosis.

[2] Peterson, Joseph, "Experiments in Rational Learning," *Psychol. Review,* 1918, 25, 462-463.

XI.

Summary and Conclusions

I. The Hard Opposites Test herein presented enables the instructor within a brief period to examine a large number of students. Even the most inexperienced will find little difficulty in evaluating the papers, so statistically simple is the test.

II. Initial experimentation began with all words previously employed but eventually the list was narrowed so as to include only fifty-four of these stimuli. Similarly the fourteen original words, for the most part of greater difficulty, were selected from a long list which was subjected to the same process of investigation.

III. Subjects were tested both orally and individually so the experimenter might study each stimulus from the standpoint of both time and accuracy. Records were kept of each individual response and reaction time. On the basis of these records the relative difficulty of the words was computed. The calculations are based on at least one hundred records for each word, varying up to one hundred and eighty-seven for others.

IV. The acceptable responses were chosen by five judges, from all those responses given by the subjects experimented upon. Additional responses which suggested themselves to each judge were recorded and in turn passed on by the others. Effort was made to have the responses include every possible opposite, not merely the most common ones. These judgments were averaged and a value of 0, ½ or 1 was accorded each response.

V. The relative difficulty of the words was determined by multiplying the median time of all correct and half-correct responses by the percentage of failures, allowing one-half a point for half-correct responses. These values were then converted into standard deviation and steps of 1/5 standard deviation were marked off. Having selected for the easiest word, a word which for college students approached zero difficulty, we assigned to it

a value of one, to the next step a value of two. The hardest word in the list received a value of twenty-six.

VI. For convenient use the test was converted into a group test. The stimuli were arranged in the ascending order of difficulty and printed on a good qualtiy of white paper. On the back of the sheet, directions make clear to both the subject and instructor what is to be done. Three thousand copies of the test were mailed to psychologists in normal schools, colleges, and universities throughout the country. From this number the records of 1628 students were obtained, on the basis of which norms for college students were established. These standards are presented in terms of percentiles for each class, including graduates, and for each sex.

VII. The list of stimuli has been divided into an equal number of words and points, for the convenience of any experimenter who may desire two lists rather than one. As it stands, the test can be easily incorporated into a group of tests and comparisons made, provided the other tests are presented in terms of percentiles. The test, because of its large number of possible points and its difficulty, offers an opportunity of testing adequately a less highly selected group than that represented by college students.

VIII. The results of this investigation tend to show that success in the test is dependent upon native ability rather than years of schooling. This conclusion seems justifiable in view of, first, the tremendous amount of overlapping among the several classes, and, second, the gradual rise of the lowest scores from year to year while the highest scores remain constant. The fact that graduates are superior to seniors, and sophomores to freshmen is to a considerable extent to be attributed to the operation of the law of selection.

APPENDIX

Name.....
School.....
Sex.....
Underscore the class of which you are a member:
Freshmen; Sophomore; Junior; Senior; Graduate.

Directions for Giving the Test

Before distributing the test blanks, request the students not to turn over the sheets until the signal is given. Ask them to fill in the blank spaces at the top of the page and then make clear what is to be done by reading aloud the following directions:

On the other side of this page are a number of words beside each of which you are to write as quickly as possible the exact opposite. For instance if the word "Black" occurs you should write White." The opposite you write must belong to the same part of speech as the word in the list. Phrases and words formed by prefixing "Non" are counted wrong. Begin at the top and work downward, but in case you do not know the opposites, pass on to the next word, and later, if you have time, come back to those you have omitted.

You will be given just six minutes, so when the Instructor says "Ready" be prepared to write, and when he says "Go" turn over the sheet and begin. Continue until the signal is given to stop.

Full	Busy	Obscure
Negative	Abstract	Proficient
After	Advance	Rigid
Dim	Foreign	Repulsion
Blunt	Create	Imaginary
Success	Simple	Permit
Pessimistic	Extravagant	Orthodoxy
Joy	Aristocratic	Analytical
Public	Rare	Extrinsic
Profit	Dangerous	Sacred
Spend	Slovenly	Dynamic
Always	Defective	Loquacious
Graceful	Stingy	Heterogeneous
Strength	Reveal	Spurious
Ancient	Diligent	Disastrous
Expand	Join	Facility
Barbarous	Impoverish	Pride
Hinder	Permanent	Result
Despondent	Elation	Adroit
Vague	Sinful	Parsimony
Fertile	Obnoxious	Suave
Doubtful	Conservative	Exoteric
Injurious	Victorious	

BIBLIOGRAPHY

BONSER, FREDERICK G., "The Reasoning Ability of Children of the Fourth, Fifth, and Sixth School Grades," *Col. Univ. Cont. Educ.*, 1906, XXXVII.

BRONNER, A. F., "A Comparative Study of the Intelligence of Delinquent Girls," *Col. Univ. Cont. Educ.*, 1914, LXVIII.

GREENE, HARRY A., "A Standardization of Certain Opposites Tests," *Jour. Educ. Psychol.*, 1918, IX, 559-566.

HERING, JOHN P., "Derivation of a Scale to Measure Abilities in Scientific Thinking," *Jour. Educ. Psychol.*, 1919, IX, 417-431.

KING, IRVING, AND GOLD, HUGO, "A Tentative Standardization of Certain Opposites Tests," *Jour. Educ. Psychol.*, 1916, VII, 459-482.

KING, IRVING, AND M'CRORY, J. L., "Freshman Tests at the State University of Iowa," *Jour. Educ. Psychol.*, 1918, IX, 32-46.

KITSON, H. D., "The Scientific Study of College Students," *Psychol. Monog.*, 1917, XXIII.

NORSWORTHY, NAOMI, "The Psychology of Mentally Deficient Children," New York, Columbia University, 1906.

PETERSON, JOSEPH, "Experiments in Rational Learning," *Psychol. Review*, 1918, XXV, 462-463.

ROSENOW, CURT, "The Analysis of Mental Functions," *Psychol. Monog.*, 1917, XXIV.

RUGER, H. A., "The Psychology of Efficiency," *Archiv. of Psychol.*, 1910, XV.

SIMPSON, B. R., "Correlations of Mental Abilities," *Col. Univ. Cont. Educ.*, 1912, LIII.

THORNDIKE, E. L., "Mental and Social Measurements," New York, 1916.

WEIDENSALL, JEAN, "The Mentality of the Criminal Woman," Baltimore, 1916.

WOODWORTH, R. S., AND WELLS, F. L., "Association Tests," *Psychol. Monog.*, 1911, LVII.

WOODY, CLIFFORD, "Measurements of Some Achievements in Arithmetic," *Col. Univ. Cont. Educ.*, 1916, LXXX.

WOOLLEY, HELEN THOMPSON, AND FISHER, CHARLOTTE RUST, "Mental and Physical Measurements of Working Children," *Psychol. Monog.*, 1914, LXXVII.